

DISSERTATIONES PHYSICAE UNIVERSITATIS TARTUENSIS

16



STUDIES OF CRYSTALLINE CELLULOSES USING POTENTIAL ENERGY CALCULATIONS

Ph. D. Thesis

by

Alvo Aabloo

TARTU 1994



STUDIES OF CRYSTALLINE CELLULOSES USING POTENTIAL ENERGY CALCULATIONS

Ph. D. Thesis

by

Alvo Aabloo

TARTU 1994

The study has been performed in University of Tartu, Institute of Experimental Physics and Technology, Tartu, Estonia.

Supervisors: Raik-Hiio Mikelsaar, Dr. Sci.
Alexander I. Pertsin, Dr. Sci.

Official Opponents: Arvi Freiberg, Dr. Sci. (Tartu)
Valery Poltev, Dr. Sci. (Puchchino)
Raivo Teeäär, Cand. Sci. (Tallinn)

The thesis will be defended on May 25, 1994 at 2 p. m. in Council Hall of University of Tartu, Ülikooli 18, EE2400 Tartu, Estonia.

Secretary of the Council:



A. Lushcik

The author was born in 1965. He graduated from University of Tartu in 1989. During 1989-1991 he worked as a junior research worker in Institute of Experimental Physics and Technology of University of Tartu. During 1991-1994 he was a PhD student at the same institute.

The permanent address of the author is:

University of Tartu, Institute of Experimental Physics and Technology, Tähe 4 Street, EE2400 Tartu, Estonia

E-mail: alvo@physic.ut.ee

© Alvo Aabloo, 1994

Table of contents

| | |
|---|----|
| 1. Introduction | |
| 1.1. Polysaccharides and cellulose as wide-spread biopolymers | 4 |
| 1.2. Survey of crystalline structure crystallographic and modelling methods of saccharides | 6 |
| 1.2.1. Experimental methods | 6 |
| 1.2.2. Theoretical methods of conformational analyses | 7 |
| 2. Methods | |
| 2.1. Molecular modelling | 9 |
| 2.2. Molecular mechanics and the minicrystal method | 11 |
| 2.2.1. Molecular mechanics | 11 |
| 2.2.2. Constructing of a minicrystal for MM3 | 12 |
| 2.3. Rigid-ring calculations - advantages and drawbacks | 13 |
| 2.4. An implementation of rigid-ring method in cellulose crystal structure refinement | 16 |
| 2.4.1. A force field | 16 |
| 2.4.2. X-ray refinement | 17 |
| 3. Results and discussion | |
| 3.1. Rigid-ring calculations improvement with different glucose rings and force fields | 19 |
| 3.2. Potential energy calculations of the crystalline structure of cellulose I | 29 |
| 3.2.1. Initial conformations | 29 |
| 3.2.2. Parallel models of cellulose I crystalline structure | 30 |
| 3.2.3. Antiparallel models of native celluloses | 31 |
| 3.3. The full molecular mechanics (MM3) calculations of celluloses | 33 |
| 3.3.1. Experimental | 33 |
| 3.3.2. Effect of the dielectric constant | 34 |
| 3.3.3. Energies of cellulose polymorphs | 35 |
| 3.4. Discussion over cellulose structure | 37 |
| 4. Conclusions | |
| 4.1. Methods | 38 |
| 4.2. Structure of celluloses | 40 |
| Acknowledgements | 41 |
| References | 41 |
| List of publications | 45 |
| Tselluloosi kristalsete faaside struktuuri uurimine kasutades energeetilisi arvutusi | 46 |

1. Introduction

1.1. Polysaccharides and cellulose as wide-spread biopolymers

Saccharides are ubiquitous in nature. They occur in all forms of life and, because of their unusual properties, present a unique source of chemicals. Saccharides of living organisms and plants perform a great biological role. They function as structural materials, energy reserves and adhesives. They appear to be essential in the process of infection by certain pathogenic species.

Cellulose is chemically a poly-(1→4)-β-D- glucopyranose - $[(C_6H_{10}O_5)_2]_n$. A cellulose chain unit consists of two pyranose rings (see Figure 1). The level of polymerization n is between 1000 and 10000, depending on the sample's nature. It is one of the most widely spread biopolymers in the world. A native sample consists of an amorphous phase as well of a crystal phase of cellulose. The latter is made up of microcrystallites. These microcrystals form fibres. This very complicated and dynamical structure makes native cellulose samples extremely flexible and strong. An investigation of the structure of this widely spread polymer seems to be important. Notwithstanding multitudinous researches carried out in the past decades, the exact structure of cellulose crystals remains without satisfactory explanation. Pure cellulose crystals exist in various forms, named I to IV, depending on the nature of the sample. Cellulose I, the native cellulose, has recently been recognized to occur as a

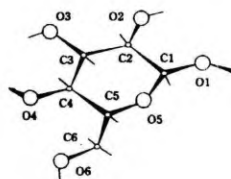


Figure 1. The pyranose ring.

compound of two polymorphs, $I\alpha$ and $I\beta$. These polymorphs occur in different ratios in different native cellulose samples. The most important industrial cellulose is cellulose II, which forms during a mercerization process (treatment in 22% sodium hydroxide) or at crystallization from solution. Cellulose III is the product of celluloses I and II treatment with liquid ammonia. Cellulose IV results from treatment in high temperature. Both phases III and IV have also two subclasses depending on their parent structure^{1,2}.

Computational methods used up to now for solving a structure of cellulose crystals have been extremely tremendous^{3,4}. They use too enormous computer resources. It is possible to refine structures with more simple algorithms. These methods will solve a structure even more precisely.

There have arisen different questions which require explanations. As the parameters of unit cells of native cellulose phases have been recently recognized⁵, the first aim of the present paper is to attempt to refine a structure of these phases. A second interesting problem dealt with in this paper is the issue of cellulose chains' direction in an unit cell. Cellulose II is considered to have an antiparallel structure⁶. It is known from experimental data that cellulose $I\alpha$ converts into cellulose $I\beta$ during annealing^{7,8} and cellulose I into cellulose II during mercerization. The question is, how the parallel cellulose I converts into the antiparallel cellulose II; whether cellulose I has an antiparallel structure or whether there exists another option.

1.2. Survey of crystalline structure crystallographic and modelling methods of saccharides

1.2.1. Experimental methods

The crystal structure analyses are generally routine; to obtain a single crystal of suitable quality may be a serious problem for the majority of saccharides. In fact, the diffraction analyses of a crystalline polymer cannot be approached in the same manner as a classical single crystal analyses. Because of the lack of diffraction data, positions of atoms cannot be determined directly from intensity data. A model analysis technique should be applied to refine the minimized differences between the experimental data and a calculated model. X-ray, electron diffraction and infrared spectroscopy⁹ are the most powerful diffraction techniques. Recent developments in solid state NMR spectroscopy, particularly the crossed polarization magic angle spinning technique indicate that this could be a very vigorous method to investigate solid state molecular conformations and environments for saccharides¹⁰. The high resolution NMR spectroscopy has become the most valuable physical implement for studying conformations of saccharides^{11,12,13}, particularly in solution. Chemical shifts, coupling constants, NOE's (Nuclear Overheimer Effect) and relaxation rates contain detailed information about the conformational structure of saccharides in solution.

1.2.2. Theoretical methods of conformational analyses

There exist various approaches to theoretical conformational analysis. The classification of these can be made in several ways. One of the possible methods is shown^{14,15} in Figure 1.

Direct methods are based on the calculations of total energy of an object which is minimized with respect to all or to some of the structural parameters. In indirect methods, on the other hand, conclusions are made on the basis of analyses of different experimental data. There are several ways to estimate the total energy of a structure in direct methods. Usually there are two or more schemes to estimate the total energy in non-uniform methods. The total energy calculations are split into different interactions. In general, there are different basis to estimate bonded and

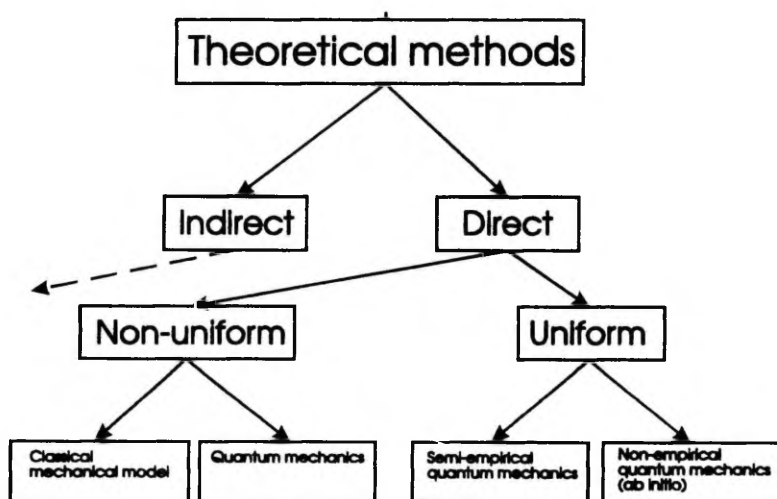


Figure 2. Classification of theoretical methods of conformational analyses.

non-bonded terms. Further, non-bonded interactions are divided into different terms. In case of a uniform method there is only one scheme to calculate total energy. This happens when applying quantum mechanical methods in which all electrons or all valence electrons are used. They constitute a group of uniform methods. With neglect of relativistic effects and within the scope of the Born-Oppenheimer approximation, the exact wave function of the structure is derived from the solutions of the Schoedinger equation¹⁸. Based on the approximations used in solving the Schoedinger equation, the uniform methods can be classified into two groups: *ab initio* (non-empirical) and semiempirical. The *ab initio* calculations need huge computer resources. Only smaller acyclic and cyclic molecules can be used as models for the structural segment studies of saccharides¹⁷. Classical methods originate from vibrational spectroscopy. The system is held together by forces which are described as potential functions of structural features, e. g. bond lengths, etc. A more detailed description of molecular mechanic methods will be given further.

There exists one mighty method of structure refinement. The latter is related to non-uniform classical methods of structure refinement and is called molecular dynamics¹⁸. It differs from other refinement implements in the sense that the aim of this method is not to minimize the total energy of the system, but to follow the dynamical state of the system. Naturally, the system moves towards its equilibrium. By application of the simulation method of MD, a trajectory (configurations a function of time) of the system can be generated by simultaneous integration of Newton's equations of motion for all atoms ($i = 1, 2, \dots, N$) in the system

$$\frac{d^2 r_i(t)}{dt^2} = m_i^{-1} F_i(t) \quad (1)$$

where the force affected on atom i with mass m_i is derived from

$$F_i(t) = - \frac{\delta V(r_1(t), r_2(t), \dots, r_N(t))}{\delta R_i(t)} \quad (2)$$

MD methods use a classical mechanical force field as a potential energy term. It means that MD cannot be more precise than the applied force field. MD, in comparison with the standard classical mechanical methods, is more powerful because it does not calculate only a static potential force field but also includes kinetic energy. The latter makes it possible to calculate dynamic states. The issue of minor local minimas is being solved. At the same time the MD requires huge computer facilities. MD is wide-spread in investigations of proteins, nucleic acids, solvations etc. It is used in structure refinement of celluloses¹⁹.

2. Methods

2.1. Molecular modelling

Despite many powerful computation methods to solve secondary and tertiary structure of biological molecules, the molecular modelling remains one of the most wide-spread methods in structure analyses. First, it has a role of visualization and demonstration of the conformation of biomolecules. Second, it is a powerful method to visualize computational methods. By using molecular modelling, we can add human thought to the refinement of biostructures. We can find the initial structures for calculations and follow the computing output. Molecular modelling can be divided into two main branches: computer graphics modelling and physical modelling.

Computer graphics modelling.

Close to computer calculation facilities. The results of calculations can be directly converted into graphics routine input and vice versa. The results of graphic modelling can serve as an input for calculations. We can also check the calculation process by monitoring intermediate results through graphic output. System modifications can be made easily. There is a lot of software built for computer graphics' modelling. By using computer graphics, we can visually control and improve several configurations of molecules. Shifting the molecules and rotating the side groups makes it possible to follow the most likely models of macromolecule systems.

*Physical molecular modelling.*²⁰

Physical models are close to a *3-dimensional reality*. Despite the fast development of graphics hardware, the computer image cannot reach the desired quality. Physical models are more handy and more informational, for they give full 3-dimensional properties. They are more convenient for demonstration and study purposes. On other the hand, the main negative qualities comprise technological difficulties of building up a perfect model and making changes in the model. Another type of negative properties concerns remoteness from computer-world. It is hard to convert the results of physical modelling for computer computation purposes and vice versa. We principally only get a basic idea from physical models on which we can build up a computer model.

2.2. Molecular mechanics and the minicrystal method

2.2.1. Molecular mechanics

Molecular mechanics (MM) as a method of conformational analyses²¹ has a wide usage being a rather simple method of substance structure refinement. We look for a minimal value of energy function. When simulating a molecular system, we postulate an energy function which describes the potential energy of the molecular system as a function of the positions r_i of the N atoms labeled by the index i . The minimizing function is the crystal's potential energy, i. e. the steric energy. The aim of this method is to find a energy minimum by changing the conformation of molecules. The MM methods²² are based on the following philosophy: a molecule is regarded as a collection of atoms held together by harmonic forces. These forces can be described by the potential functions of structural features. The main feature is to use a simplified parametric force field instead of solving complicated equations. The need of parametrization results from the enormous amount of calculations needed to solve the conformation of a molecule. Even semiempirical valence-electron approximation¹⁶ calculations are too extensive to solve the structure of bigger molecules, not do mention the biological macromolecule and crystal structures. All of the most widely spread force fields use a bond-related ideology - parameters are associated with bonds, bond angles, torsional angles or distances between two atoms. The empirical functions have been suggested in several works^{23,24}. Depending on requirements they can be more complicated with more correction members. The values of parameters and formulas of empirical functions are derived from *ab initio* calculations, semiempirical calculations and experimental data on the conformations. We used the MM3(90)

program^{25,26} for full molecular mechanics^{27,28} calculations. A thorough description of this program can be found elsewhere^{29,30}.

2.2.2. Constructing of a minicrystal for MM3

The MM3 program is able to use a maximum of 800 atoms with block diagonal minimization option standard. This limits the construction of crystals. With the number of atoms increased, the minimization requires more time. Each atom adds six more degrees of freedom into conformational space. As MM3 does not have any crystal border effect facilities, it is necessary to construct a minicrystal¹. In order to establish a crystal force field for cellulose chains. For this reason we built in the course of the research cellulose chains as cellotetroses (see Figure 3) and arranged seven cellotetroses into crystal packing¹ in accordance with unit cell parameters (see Figure I.2 and Figure IV.2) refined from experimental diffraction data. We presumed that at this minicrystal case glucose rings that are situated in the central part of a minicrystal should have an average and periodical force field.

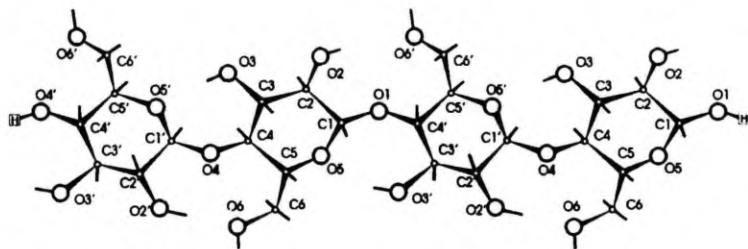


Figure 3. A cellotetrose of a cellulose chain used in minicrystals.

This approximation is relatively good as most of the interactions vanished significantly at distances less than 1 nm. We also added terminating hydrogen atoms in the ends of cellotetroses. This is needed for neutralizing charges that we used for the calculations of electrostatic interaction. The MM3 routine does not use any cutoff distances in atom-atom interaction calculations. VanderWaal's and electrostatic interactions will be calculated over all atoms. The orientation of terminal hydroxyl groups was random. Their positions optimized during minimization.

2.3. Rigid-ring calculations - advantages and drawbacks

Full molecular mechanics calculates potential energy of structures (crystal) including all components of a force field. In theory, we should get a heat of formation of crystal as a result. As it was already mentioned above, conformation of molecules is defined by the following interactions: bond lengths, bond angles, torsion angles, non-bonded (VanderWaal's and electrostatic), and hydrogen bond interactions. Variations in the molecular geometry of molecules are then very simply defined as changes in bond length, bond angle or torsional angle. Application of a typical force constant of bond stretching²⁸ and assuming Hook's law dependence indicate, that the distortion of a single bond of 0.03 Å would cost about 1.2 kJ/mol. A bond angle bending is less sensitive, and a bond stretch about 0.05 Å is equal to an angular distortion of about 10° ¹⁵. Torsional changes involve rotation around bond axis. The barrier to rotation around a single C-C bond is 12.3 kJ/mol. The barrier to rotation of methoxyl group in dimethoxymethane is approximately 4.2 kJ/mol²⁹. At the same time distortion of hydrogen bond of 0.1 Å costs less

than 0.3 kcal/mol. These values show that different terms in the function of potential energy have different "sensitivity". Sometimes there is no need to calculate full potential energy for solving structures. The structure is determined simply by some components of force field. Other terms are not changing remarkably and only disturbing a minimization process. The crystal structure of celluloses is mostly determined by hydrogen bond and non-bonded interactions between the chains of cellulose. Due to stronger interactions (bond lengths and bond angles of atoms in sugar ring) the geometry of glucose rings is rigid enough. The values of these conformational parameters can be reached by an experiment or by *ab initio* and semiempirical calculations of similar and more simple compounds. This is called a linked atom approach³¹. There are two reasons why the above-mentioned terms disturb structure refinement.

First, by excluding these interactions from minimizing functions, we simply decrease the number of variables. By this we make the refinement algorithm more effective. Effectiveness depends on the algorithm we use.

Second, these components may be extremely "strong" in comparison with others (see page 14). It means that small changes in conformation cause relatively higher energy changes from *equilibrium* state. We have a situation where some of the components of a force field are significantly more intensive than others. The minimizing routine traps because of these interactions. It starts to oscillate around the equilibrium states of these components and, minimizing other interactions, remains on the background. On the other hand, these terms are playing a leading role in the formation of the crystals of polysaccharides. If we turn the terms which are more efficient into constants, it is also possible to refine the terms that are weaker but have an important role in structure formation.

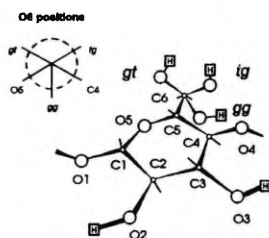
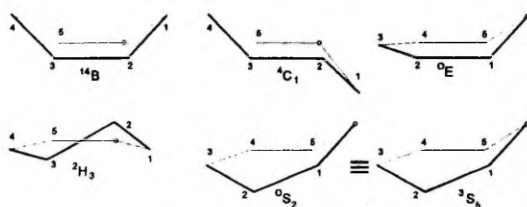


Figure 4. Different O 6 positions (*tg, gg* and *gt*) in glucose ring.

This method has many dangerous nuances. When fixing some conformational parameters as bond lengths or bond angles, we must be assured that these parameters would not change remarkably. Moreover, small deviances from the best solution - in force field meaning - do not significantly affect the results of refinement. Even if we take these values from similar substances, they are not exactly the same as in our structure. The geometry of

molecules in crystal structure can be different from their equilibrium in a free state, e. g. in solution. The final energy of all crystal structures can even be lower if a molecule has been distorted. For example, it is known from experimental data and theoretical calculations that hydroxymethyl groups of polysaccharide chains are preferable in *gg* position (see Figure 4). However, as it will be shown later, it seems that in celluloses they are preferably in *tg* position. At the same time, a glucose ring seems to be extremely stable. Calculations with MM3 show³² in 99.99% percent of the cases that a pyranose ring is in the ⁴C₁ conformation position (see Figure 5).



There are 36 characteristic conformations of pyranose ring, including 6 boats (B), 2 chairs (C), 12 envelopes (E), 12 half - chairs (H) and 6 skewes (S). Other combinations are redundant, as shown at figure.

Figure 5. Some examples of different conformations of a pyranose ring.

2.4. An implementation of rigid-ring method in cellulose crystal structure refinement

2.4.1. A force field

In the following discussion of structure refinement we will apply the rigid-ring refinement method to find the crystal structure of different polymorphs of cellulose. As the structure of native celluloses is not clearly solved yet, this is a good reason to try to improve different models for these phases by rigid-ring calculations. The method to build up cellulose chains from glucose rings using virtual bond³³ and a unit cell from these chains is described elsewhere^{1,34}. Measurements of unit cells of native celluloses is taken from *Sugiyama et al*⁶. Experimental data for cellulose II is taken from *Kolpak et al*²⁵.

All calculations have been made in internal coordinates - bond lengths, bond angles etc. - and the program uses the so-called Z-matrix coordinates. As a force field we have used a simple one without an outlined electrostatic interaction term. The electrostatic interaction has been considered in other terms. Avoiding direct electrostatic interaction simplifies the usage of periodic boundary conditions and an infinite chain in calculations. As we did not vary any bond length, this interaction term is not necessary.

Bond angle bending potential. In equation (3) θ is a bond angle, θ_0 is a equilibrium bond angle and k_θ is a force constant.

$$E_\theta = k_\theta (\theta - \theta_0)^2 \quad (3)$$

Torsional potential.³⁶ In equation (4) ω is a torsional angle U_0 is a force

constant.

$$E_{\omega} = \frac{U_0}{2}(1 + \cos 3\omega) \quad (4)$$

Hydrogen bond potential.^{36,37} A Morse equation was used for hydrogen bond term calculations. In the equation (4) r is a distance between the acceptor oxygen and donor hydrogen atoms. D is a force constant.

$$E_H = D[e^{-6(r-r_0)} - 2e^{-3(r-r_0)}] \quad (5)$$

Non-bonded atom-atom potential.³⁸ A Buckingham expression was used for non-bonded interaction modelling. In the equation (6) r is a distance between atoms, A , B and C are force constants. As this term includes VanderWaals', electrostatic and other interactions between two non-bonded atoms, we use different force constants in case of intramolecular interactions and intermolecular interactions³⁹.

$$E_{nb} = -Ar^6 + Be^{-Cr} \quad (6)$$

This force field is simple compared to the MM3's field. As we do not vary the glucose ring or any bond length, these terms are sufficient enough to express forces in our case.

2.4.2. X-ray refinement

The difficulties with cellulose are not unexpected regarding its comparatively poor quality of diffraction pattern from oriented cellulose samples. Typical X-ray diagrams of cellulose II, III and IV contain only a few dozens reflections, which is clearly insufficient to refine all atomic parameters by standard crystallographic

studies. X-ray data of cellulose I is not available yet, as the phase I α coexists only with the phase I β . To increase the terms of refinement of the cellulose crystalline structure, we use X-ray diffraction patterns as part of the minimizing function. We compute the objective function

$$\Phi = U + WR'' \quad (7)$$

where U is a potential energy, W - a weighing factor, R'' - a crystallographic discrepancy factor.

R'' -factor is defined by

$$R'' = \left(\frac{\sum_{m=1}^M \omega_m |F_m^{obs} - F_m^{calc}|^2}{\sum_{m=1}^M \omega_m |F_m^{obs}|^2} \right)^{\frac{1}{2}} \quad (8)$$

where F_m^{obs} and F_m^{calc} are the observed and calculated structure factor amplitudes, ω_m is the weight factor applied to the m -th reflection. M is the number of observed reflections. Each F_m^{calc} is a function of the parameters of model and is computed from:

$$F_m^{calc} = K \left\{ \sum_{hkl} [F_{hkl}^{calc} e^{-\frac{B \rho_m^2}{4}}]^2 \right\}^{\frac{1}{2}} \quad (9)$$

The summation in this equation is over all planes hkl , contributing to the m -th reflection, ρ_m is the reciprocal d -spacing, K is the scale factor.

We can see from the equation (7) that the calculated R'' -factor is given a weight of W . The value of W is chosen in order as to make small changes in R'' and U equal meaning for objective function. The idea is that statistically significant changes in R'' -factor should be equal to significant changes in potential energy

level. A significant level can be obtained from Hamilton' tables⁴⁰. The latter depends on a number of variables. In potential energy calculations the typical level of accuracy is about 1 kcal/mol⁴¹.

As mentioned before, we could use this objective function refinement only for the crystal structure of cellulose II. It is not possible in case of native celluloses the X-ray data and we can refine the structure by using potential energy minimization.

3. Results and discussion

3.1. Rigid-ring calculations improvement with different glucose rings and force fields

In rigid-ring calculations we used fixed glucose rings and made an attempt to investigate the influence of residue geometry on the results of minimization¹¹. Using different glucose rings, we found the minimums for several crystalline models of cellulose II described elsewhere⁴¹. We improved six different glucose ring conformations^{42,43}, including even α -glucose rings which were converted into β -glucose by chiral inversion of residues. All other conformational parameters of minimization were kept the same. Similarly, several minimization technique were used to obtain the best results. All results are presented elsewhere¹¹. Tables 1 and 2 show the objective function and the potential energies of the thirteen most probable models of crystalline structure of cellulose II. Consequently, we can see that the best and most important models do not change remarkably. Only one model (A11) was not found in some cases and it was minimized to model A1. In

Table 3 there are presented the mean movements of the models according to Arnott-Scott average ring case. We can see that generally all results of different rings for most models have good accordance with Arnott-Scott ring results. It seems that models with lower energy deviate less during the change of the glucose ring. In Table 4 are presented variable torsion angles as the results of different rigid-ring minimizations with different glucose rings.

Another attempt to improve the rigid-ring method was made. We built an alternative force field²³. As this force field includes an electrostatic equation, due to boundary condition, some virtual hydrogens were included in the boundary glucose rings to correct the charge neutrality problem. Positions of these hydrogens were not minimized. We used an electrostatic constant of 4^{LVH} as in the case of full minimization technique of MM3. Results of these calculations are shown in Table 5, 6 and 7. We can see that in this case the best models remain in their present positions. The values of energies are different. There exist different reasons for the diverse energy values obtained as a result of crystal energy minimizations are different. The most important of them is that during fixed-ring the minimization we switched off some interactions. If we change the force field, the components of force field act in a different way. The value of crystal energy obtained by the minimization process has minor significant meaning. The analyze is possible by comparing different values and making conclusions at this level. We can see that these results are somewhat dispersed in comparison with different ring calculations, but the best models are in their place. The three last models (A10-A12) were not found in some cases and were minimized to positions of other models. In Table 6 the mean movement was calculated against the conformations that these models had in the initial force field. Similarly to the previous computer experiment, the models which had lower energy terms deviated less during different minimization processes. We can see that values of the mean atomic

movement are low.

According to these calculation experiments we can say that small changes in glucose rings do not remarkably influence the minimization results. Also, having used different force fields that gave us similar results in full minimization process⁴, we can conclude that all rigid-ring calculations gave the same kind of structures. We also saw that structures with lower crystal energy have a lower mean atom movement in different conditions. As the model with the lowest energy is also the most probable model, it depends least of all on minor changes in a glucose ring or on force field deviations. The glucose ring is rigid enough in polysaccharides to omit its degrees of freedom in a energy term function. Components of force field related to bond lengths, bond angles and glucose ring conformations are too efficient in comparison to the components related to torsional angles of side groups and hydrogen bonds. These forces play a leading role in the formation of crystal phases of polysaccharides. By decreasing the amount of degrees of freedom in a minimizing function, we make the minimization process more efficient as we decrease the amount of variables in potential energy function from 285 to 12 (in case of phase Ia).

Table 1. Objective functions of most probable models of cellulose II crystals using different rigid glucose rings. Glucose rings: AS - average Arnott-Scott; CHA2 - cyclohexaamylose; PLA - planteose; GUR - glucose-urea; CBIO - cellobiose; β GLU - β -D-glucose. A1,..., A12, P1 - most probable models of the unit cell of cellulose II.

| Models | Objective functions with different glucose rings (kcal/mol) | | | | | |
|--------|---|--------|--------|--------|--------|--------|
| | AS | CHA2 | PLA | GUR | CBIO | BGLU |
| A1 | -18.70 | -19.75 | -18.53 | -19.05 | -19.94 | -18.60 |
| A2 | -18.50 | -19.31 | -19.25 | -18.41 | -19.72 | -18.41 |
| A3 | -18.54 | -18.00 | -17.90 | -17.89 | -18.60 | -19.03 |
| A4 | -17.59 | -16.88 | -15.98 | -17.49 | -16.97 | -18.70 |
| A5 | -17.60 | -19.10 | -18.13 | -17.22 | -18.99 | -17.34 |
| A6 | -17.48 | -16.90 | -18.37 | -18.60 | -17.21 | -18.15 |
| A7 | -17.40 | -16.41 | -16.53 | -17.83 | -17.34 | -18.06 |
| P1 | -16.90 | -18.00 | -16.67 | -15.95 | -19.32 | -16.43 |
| A8 | -16.40 | -15.86 | -17.00 | -16.19 | -16.61 | -16.80 |
| A9 | -15.80 | -14.99 | -16.41 | -16.83 | -15.43 | -16.63 |
| A10 | -15.10 | -14.20 | -10.03 | -12.76 | -9.11 | -13.01 |
| A11 | -8.00 | -13.10 | -11.50 | -9.54 | -19.80 | -18.64 |
| A12 | -7.80 | -8.10 | -11.83 | -14.48 | -15.20 | -15.11 |

Table 2. Potential energies of most probable models of the cellulose II crystal using different fixed glucose rings. AS - average Arnott-Scott; CHA2 - cyclohexaamylose; PLA - planteose; GUR - glucose-urea; CBIO - cellobiose; β GLU - β -D-glucose. A1,..., A12, P1 - most probable models of the unit cell of cellulose II.

| Models | Energies with different glucose rings (kcal/mol) | | | | | |
|--------|--|--------|--------|--------|--------|--------|
| | AS | CHA2 | PLA | GUR | CBIO | BGLU |
| A1 | -21.40 | -22.62 | -21.38 | -21.86 | -22.89 | -21.39 |
| A2 | -21.20 | -22.04 | -21.94 | -21.49 | -22.01 | -21.40 |
| A3 | -21.20 | -20.67 | -20.66 | -20.94 | -21.22 | -21.89 |
| A4 | -20.10 | -20.06 | -18.91 | -20.10 | -19.98 | -21.89 |
| A5 | -20.39 | -21.80 | -20.67 | -20.39 | -21.52 | -20.20 |
| A6 | -19.40 | -19.33 | -20.49 | -20.54 | -19.84 | -20.57 |
| A7 | -20.50 | -19.40 | -20.22 | -21.01 | -19.82 | -21.21 |
| P1 | -19.40 | -21.20 | -19.31 | -18.52 | -18.17 | -22.07 |
| A8 | -18.30 | -18.07 | -15.86 | -18.34 | -19.07 | -18.75 |
| A9 | -17.60 | -17.32 | -18.38 | -18.63 | -17.68 | -18.54 |
| A10 | -17.10 | -14.50 | -13.38 | -15.28 | -11.94 | -15.21 |
| A11 | -9.70 | -13.30 | -13.51 | -11.36 | -22.63 | -21.47 |
| A12 | -9.40 | -14.70 | -13.87 | -16.51 | -17.81 | -18.11 |

Table 3. The mean atomic movements that result from the change of glucose ring at rigid-ring minimization of cellulose II crystals. For abbreviations see Table 2. The initial glucose ring was the average Arnott-Scott ring.

| Models | Mean movements | | | | |
|--------|----------------|-------|-------|-------|-------|
| | CHA2 | PLA | GUR | CBIO | BGLU |
| A1 | 0.040 | 0.093 | 0.081 | 0.172 | 0.110 |
| A2 | 0.057 | 0.112 | 0.109 | 0.157 | 0.171 |
| A3 | 0.102 | 0.054 | 0.067 | 0.108 | 0.146 |
| A4 | 0.101 | 0.186 | 0.101 | 0.011 | 0.090 |
| A5 | 0.093 | 0.148 | 0.092 | 0.185 | 0.059 |
| A6 | 0.025 | 0.135 | 0.190 | 0.101 | 0.168 |
| A7 | 0.086 | 0.078 | 0.108 | 0.169 | 0.086 |
| P1 | 0.183 | 0.074 | 0.201 | 0.139 | 0.113 |
| A8 | 0.139 | 0.183 | 0.103 | 0.137 | 0.107 |
| A9 | 0.131 | 0.203 | 0.139 | 0.184 | 0.118 |
| A10 | 0.109 | 0.056 | 0.087 | 0.264 | 0.092 |
| A11 | 0.299 | 0.164 | 0.179 | 0.214 | 0.320 |
| A12 | 0.241 | 0.203 | 0.181 | 0.213 | 0.248 |

Table 4. Variable torsion angles of different cellulose II crystal models using several glucose rings. For abbreviation see Table 2.

| Models | Glucose rings | τ_{11} | τ_{12} | τ_{13} | τ_{14} | τ_{21} | τ_{22} | τ_{23} | τ_{24} |
|--------|---------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| | | | | | | | | | |
| A1 | CHA2 | 73 | -48 | -151 | 167 | 74 | 164 | -54 | 157 |
| | PLA | 66 | -51 | -151 | 174 | 69 | 162 | -53 | 163 |
| | GUR | 66 | -51 | -151 | 174 | 69 | 162 | -53 | 162 |
| | BGLU | 72 | -56 | -167 | 181 | 71 | 178 | -58 | 184 |
| | CBIO | 68 | -51 | -147 | 190 | 74 | 181 | -59 | 167 |
| | AS | 66 | -51 | -151 | 174 | 69 | 162 | -53 | 163 |
| A2 | CHA2 | 73 | -48 | -152 | 166 | 76 | 103 | -56 | 158 |
| | PLA | 68 | -51 | -152 | 175 | 70 | 103 | -57 | 163 |
| | GUR | 68 | -51 | -152 | 176 | 71 | 103 | -57 | 163 |
| | BGLU | 75 | -52 | -152 | -197 | 76 | 108 | -56 | 167 |
| | CBIO | 73 | -55 | -150 | -197 | 70 | 115 | -62 | 183 |
| | AS | 68 | -52 | -152 | -176 | 71 | 103 | -57 | 163 |
| A3 | CHA2 | 71 | 58 | -80 | 175 | 74 | 167 | -53 | 173 |
| | PLA | 65 | 57 | -79 | 177 | 71 | 166 | -52 | 175 |
| | GUR | 65 | 57 | -79 | 177 | 71 | 166 | -52 | 175 |
| | BGLU | 65 | 60 | -84 | 189 | 72 | 163 | -57 | 180 |
| | CBIO | 72 | 60 | -80 | 189 | 73 | 173 | -56 | 191 |
| | AS | 65 | 57 | -79 | 177 | 71 | 166 | -52 | 175 |
| A4 | CHA2 | -69 | 176 | -163 | 173 | 72 | 164 | -54 | 158 |
| | PLA | -63 | 173 | -161 | 182 | 69 | 163 | -52 | 174 |
| | GUR | -63 | 173 | -161 | 182 | 69 | 163 | -52 | 174 |
| | BGLU | -63 | 186 | -158 | 182 | 69 | 167 | -56 | 191 |
| | CBIO | -70 | 193 | -164 | 181 | 74 | 165 | -52 | 174 |
| | AS | -63 | 173 | -161 | 182 | 69 | 163 | -52 | 174 |
| A5 | CHA2 | 71 | -48 | -152 | -54 | 73 | 101 | -56 | 158 |
| | PLA | 66 | -50 | -152 | -55 | 68 | 102 | -56 | 160 |
| | GUR | 67 | -50 | -152 | -55 | 68 | 102 | -56 | 161 |
| | BGLU | 69 | -53 | -166 | -56 | 71 | 99 | -59 | 163 |
| | CBIO | 71 | -53 | -169 | -58 | 72 | 106 | -60 | 171 |
| | AS | 67 | -50 | -152 | -55 | 68 | 102 | -56 | 161 |
| A6 | CHA2 | 177 | 165 | -97 | 172 | 75 | 161 | -57 | 166 |
| | PLA | 179 | 164 | -96 | 164 | 70 | 159 | -55 | 164 |
| | GUR | 173 | 164 | -97 | 174 | 72 | 161 | -56 | 167 |
| | BGLU | 199 | 180 | -107 | 179 | 80 | 177 | -56 | 173 |
| | CBIO | 198 | 182 | -95 | 172 | 69 | 180 | -59 | 176 |
| | AS | 178 | 164 | -97 | 174 | 71 | 161 | -56 | 167 |

Table 4. (continuing)

| Models | Glucose rings | τ_{11} | τ_{12} | τ_{13} | τ_{14} | τ_{21} | τ_{22} | τ_{23} | τ_{24} |
|--------|---------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| A7 | CHA2 | 70 | -51 | -153 | 176 | 74 | 35 | -54 | 72 |
| | PLA | 67 | -52 | 153 | 176 | 70 | 34 | -54 | 72 |
| | GUR | 65 | -52 | -153 | 175 | 70 | 34 | -54 | 71 |
| | BGLU | 67 | -51 | -161 | 191 | 70 | 33 | -52 | 79 |
| | CBIO | 69 | -53 | -149 | 177 | 75 | 35 | -59 | 73 |
| | AS | 67 | -52 | -153 | 176 | 70 | 34 | -54 | 72 |
| P1 | CHA2 | 72 | 64 | -68 | 163 | 177 | 174 | -78 | 161 |
| | PLA | 64 | 64 | -70 | 172 | 171 | 175 | -77 | 169 |
| | GUR | 64 | 64 | -70 | 172 | 171 | 175 | -77 | 169 |
| | BGLU | 63 | 68 | -78 | 171 | 181 | 175 | -77 | 180 |
| | CBIO | 62 | 71 | -76 | 171 | 165 | 185 | -81 | 176 |
| | AS | 64 | 64 | -70 | 172 | 170 | 175 | -77 | 169 |
| A8 | CHA2 | 174 | 166 | -124 | 172 | 75 | 61 | 176 | 166 |
| | PLA | 174 | 166 | -123 | 177 | 64 | -65 | 177 | 166 |
| | GUR | 174 | 166 | -123 | 177 | 64 | -65 | 177 | 166 |
| | BGLU | 195 | 171 | -131 | 185 | 66 | -65 | 193 | 181 |
| | CBIO | 186 | 164 | -121 | 198 | 65 | -72 | 183 | 172 |
| | AS | 175 | 166 | -123 | 177 | 64 | -65 | 177 | 166 |
| A9 | CHA2 | 178 | 166 | -110 | 170 | 81 | 171 | 181 | 166 |
| | PLA | 177 | 162 | -109 | 175 | 77 | 172 | 183 | 167 |
| | GUR | 177 | 162 | -109 | 176 | 76 | 172 | 183 | 167 |
| | BGLU | 182 | 173 | -115 | 188 | 79 | 177 | 195 | 178 |
| | CBIO | 191 | 176 | -107 | 172 | 81 | 171 | 201 | 176 |
| | AS | 177 | 162 | -109 | 176 | 76 | 172 | 183 | 167 |
| A10 | CHA2 | -178 | 208 | 86 | 173 | -171 | -57 | -30 | 166 |
| | PLA | -199 | 204 | 87 | 174 | -170 | -57 | -28 | 170 |
| | GUR | -178 | 187 | 88 | 167 | -155 | -54 | -27 | 166 |
| | BGLU | -193 | 175 | 93 | 188 | -161 | -69 | -36 | 183 |
| | CBIO | -192 | 180 | 90 | 192 | -171 | -63 | -35 | 186 |
| | AS | -178 | 168 | 88 | 174 | -164 | -62 | -35 | 167 |
| A11 | CHA2 | 77 | -52 | -164 | 183 | 103 | 171 | -55 | 177 |
| | PLA | 84 | -54 | -157 | 176 | 109 | 174 | -58 | 170 |
| | GUR | 77 | -48 | -156 | 173 | 102 | 162 | -52 | 160 |
| | BGLU | 68 | -52 | -161 | 189 | 68 | 167 | -55 | 166 |
| | CBIO | 66 | -52 | -152 | 181 | 70 | 161 | -52 | 159 |
| | AS | 77 | -48 | -156 | 173 | 102 | 162 | -52 | 160 |

Table 5. Potential energies of most probable models of the cellulose II crystal using different glucose ring and an alternative force field. For abbreviations see Table 2.

| Models | Potential energies(kcal/mol) | | | | | |
|--------|------------------------------|--------|--------|--------|--------|--------|
| | AS | CHA2 | PLA | GUR | CBIO | BGLU |
| A1 | -22.01 | -23.12 | -21.91 | -22.70 | -23.79 | -22.58 |
| A2 | -21.95 | -22.32 | -22.80 | -21.92 | -22.26 | -21.25 |
| A3 | -21.20 | -20.82 | -21.24 | -21.49 | -21.12 | -21.54 |
| A4 | -20.67 | -19.14 | -19.89 | -19.70 | -19.17 | -21.26 |
| A5 | -20.13 | -19.45 | -20.76 | -19.58 | -20.57 | -20.97 |
| A6 | -19.80 | -19.19 | -20.68 | -20.78 | -19.17 | -20.78 |
| A7 | -19.30 | -19.58 | -20.08 | -19.39 | -18.92 | -20.22 |
| P1 | -18.93 | -19.20 | -19.72 | -18.86 | -18.28 | -18.07 |
| A8 | -17.83 | -17.16 | -16.69 | -18.95 | -21.36 | -19.23 |
| A9 | -18.34 | -17.73 | -17.81 | -18.85 | -19.69 | -18.08 |
| A10 | -16.51 | -15.39 | -12.92 | -15.39 | -17.15 | -17.77 |
| A11 | -8.90 | -14.05 | -14.49 | -11.52 | -11.71 | -15.35 |
| A12 | -8.82 | -14.04 | -13.56 | -17.39 | -22.73 | -20.34 |

Table 6. The mean atomic movements that result from the change of force field at rigid-ring minimization of cellulose II crystals. For abbreviations see Table 2.

| Models | Mean movements | | | | | |
|--------|----------------|-------|-------|-------|-------|-------|
| | AS | CHA2 | PLA | GUR | CBIO | BGLU |
| A1 | 0.008 | 0.067 | 0.127 | 0.092 | 0.103 | 0.131 |
| A2 | 0.010 | 0.076 | 0.295 | 0.147 | 0.092 | 0.077 |
| A3 | 0.006 | 0.114 | 0.269 | 0.184 | 0.219 | 0.124 |
| A4 | 0.016 | 0.066 | 0.276 | 0.090 | 0.195 | 0.230 |
| A5 | 0.012 | 0.147 | 0.074 | 0.047 | 0.175 | 0.112 |
| A6 | 0.019 | 0.128 | 0.283 | 0.211 | 0.195 | 0.139 |
| A7 | 0.007 | 0.059 | 0.221 | 0.124 | 0.250 | 0.153 |
| P1 | 0.004 | 0.086 | 0.186 | 0.127 | 0.053 | 0.105 |
| A8 | 0.008 | 0.063 | 0.293 | 0.070 | 0.119 | 0.203 |
| A9 | 0.009 | 0.114 | 0.213 | 0.083 | 0.205 | 0.075 |
| A10 | 0.019 | 0.101 | 0.259 | 0.159 | 0.226 | 0.134 |
| A11 | 0.018 | 0.081 | 0.289 | 0.096 | 0.166 | 0.083 |
| A12 | 0.007 | 0.086 | 0.125 | 0.098 | 0.128 | 0.165 |

Table 7. Variable torsion angles of different cellulose II crystals models using several glucose rings and an alternative force field. For abbreviations see Table 2 and Figure X.2.

| Models | Glucose rings | τ_{11} | τ_{12} | τ_{13} | τ_{14} | τ_{21} | τ_{22} | τ_{23} | τ_{24} |
|--------|---------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| A1 | CHA2 | 73 | -48 | -151 | 163 | 72 | 168 | -53 | 156 |
| | PLA | 66 | -50 | -149 | 173 | 67 | 158 | -54 | 160 |
| | GUR | 67 | -50 | -152 | 172 | 67 | 157 | -52 | 162 |
| | BGLU | 70 | -55 | -164 | 179 | 70 | 178 | -59 | 160 |
| | CBIO | 68 | -50 | -148 | 193 | 72 | 182 | -59 | 165 |
| | AS | 66 | -52 | -152 | 178 | 69 | 163 | -52 | 166 |
| A2 | CHA2 | 72 | -48 | -149 | 167 | 78 | 104 | -54 | 160 |
| | PLA | 68 | -50 | -150 | 173 | 68 | 100 | -58 | 160 |
| | GUR | 68 | -52 | -156 | 171 | 71 | 102 | -55 | 168 |
| | BGLU | 76 | -53 | -150 | -193 | 75 | 107 | -55 | 167 |
| | CBIO | 74 | -54 | -153 | -193 | 69 | 113 | -64 | 183 |
| | AS | 70 | -53 | -154 | -172 | 70 | 104 | -59 | 159 |
| A3 | CHA2 | 71 | 59 | -81 | 179 | 76 | 168 | -53 | 170 |
| | PLA | 64 | 57 | -80 | 179 | 72 | 164 | -51 | 179 |
| | GUR | 63 | 58 | -78 | 180 | 73 | 169 | -52 | 178 |
| | BGLU | 65 | 60 | -82 | 193 | 73 | 163 | -59 | 185 |
| | CBIO | 73 | 60 | -80 | 189 | 72 | 176 | -56 | 196 |
| | AS | 65 | 58 | -80 | 174 | 71 | 170 | -53 | 174 |
| A4 | CHA2 | -69 | 171 | -158 | 175 | 71 | 168 | -53 | 154 |
| | PLA | -61 | 175 | -164 | 177 | 69 | 159 | -51 | 176 |
| | GUR | -63 | 176 | -163 | 185 | 71 | 162 | -52 | 174 |
| | BGLU | -64 | 189 | -162 | 188 | 69 | 163 | -54 | 192 |
| | CBIO | -70 | 195 | -161 | 180 | 75 | 162 | -52 | 175 |
| | AS | -63 | 177 | -162 | 187 | 67 | 163 | -51 | 174 |
| A5 | CHA2 | 69 | -48 | -155 | -54 | 74 | 99 | -57 | 158 |
| | PLA | 66 | -50 | -148 | -55 | 68 | 99 | -56 | 156 |
| | GUR | 69 | -50 | -147 | -54 | 69 | 102 | -57 | 157 |
| | BGLU | 71 | -54 | -165 | -55 | 71 | 97 | -58 | 165 |
| | CBIO | 70 | -54 | -171 | -56 | 70 | 106 | -60 | 167 |
| | AS | 68 | -51 | -149 | -56 | 68 | 100 | -58 | 164 |
| A6 | CHA2 | 172 | 164 | -97 | 170 | 78 | 162 | -57 | 164 |
| | PLA | 175 | 166 | -95 | 168 | 70 | 162 | -55 | 166 |
| | GUR | 172 | 167 | -96 | 177 | 72 | 158 | -56 | 168 |
| | BGLU | 193 | 178 | -109 | 180 | 78 | 178 | -58 | 172 |
| | CBIO | 196 | 181 | -94 | 170 | 71 | 182 | -60 | 181 |
| | AS | 180 | 168 | -97 | 170 | 73 | 163 | -56 | 165 |

Table 7. (continuing)

| Models | Glucose rings | τ_{11} | τ_{12} | τ_{13} | τ_{14} | τ_{21} | τ_{22} | τ_{23} | τ_{24} |
|--------|---------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| A7 | CHA2 | 71 | -51 | -150 | 177 | 75 | 34 | -56 | 71 |
| | PLA | 68 | -52 | 157 | 179 | 66 | 34 | -54 | 71 |
| | GUR | 66 | -53 | -157 | 178 | 70 | 34 | -55 | 72 |
| | BGLU | 66 | -52 | -159 | 188 | 70 | 32 | -52 | 81 |
| | CBIO | 69 | -53 | -148 | 174 | 74 | 34 | -59 | 73 |
| | AS | 68 | -53 | -154 | 173 | 69 | 33 | -54 | 73 |
| P1 | CHA2 | 74 | 62 | -67 | 161 | 176 | 177 | -76 | 160 |
| | PLA | 65 | 62 | -68 | 169 | 167 | 173 | -75 | 164 |
| | GUR | 66 | 65 | -71 | 174 | 172 | 175 | -79 | 169 |
| | BGLU | 64 | 69 | -79 | 168 | 177 | 177 | -75 | 178 |
| | CBIO | 63 | 73 | -76 | 175 | 169 | 190 | -81 | 172 |
| | AS | 63 | 64 | -69 | 173 | 170 | 172 | -75 | 166 |
| A8 | CHA2 | 176 | 166 | -122 | 174 | 76 | 62 | 179 | 167 |
| | PLA | 175 | 164 | -121 | 180 | 63 | -65 | 179 | 164 |
| | GUR | 175 | 162 | -123 | 173 | 66 | -63 | 173 | 163 |
| | BGLU | 193 | 171 | -128 | 180 | 67 | -63 | 189 | 186 |
| | CBIO | 184 | 160 | -122 | 192 | 67 | -73 | 180 | 167 |
| | AS | 172 | 161 | -120 | 181 | 66 | -64 | 177 | 161 |
| A9 | CHA2 | 183 | 161 | -108 | 166 | 80 | 167 | 177 | 164 |
| | PLA | 182 | 163 | -110 | 170 | 77 | 171 | 188 | 167 |
| | GUR | 176 | 158 | -112 | 171 | 76 | 175 | 186 | 167 |
| | BGLU | 182 | 169 | -114 | 192 | 77 | 175 | 197 | 180 |
| | CBIO | 189 | 171 | -105 | 172 | 80 | 173 | 197 | 174 |
| | AS | 179 | 162 | -111 | 180 | 77 | 168 | 183 | 164 |
| A10 | CHA2 | -173 | 212 | 88 | 175 | -172 | -58 | -30 | 168 |
| | PLA | -194 | 202 | 89 | 172 | -166 | -57 | -28 | 175 |
| | GUR | -162 | 172 | 76 | 218 | -173 | -60 | -32 | 181 |
| | BGLU | -148 | 161 | 78 | 201 | -161 | -58 | -32 | 183 |
| | CBIO | -160 | 165 | 80 | 191 | -179 | -63 | -32 | 179 |
| | AS | -174 | 173 | 86 | 179 | -166 | -63 | -36 | 169 |
| A11 | CHA2 | 74 | -52 | -162 | 187 | 104 | 174 | -56 | 174 |
| | PLA | 83 | -54 | -159 | 176 | 109 | 177 | -57 | 167 |
| | GUR | 79 | -48 | -154 | 177 | 99 | 165 | -52 | 163 |
| | BGLU | 71 | -44 | -126 | 195 | 104 | 153 | -60 | 152 |
| | CBIO | 78 | -46 | -137 | 191 | 103 | 148 | -57 | 163 |
| | AS | 78 | -48 | -151 | 178 | 104 | 164 | -52 | 156 |

3.2. Potential energy calculations of the crystalline structure of cellulose I

3.2.1. Initial conformations

Besides the problem of a correct force field another major issue in the methods based on molecular mechanics is the problem concerning initial models and local minimas. As we already discussed, in case of MM methods, when calculating only potential energy of crystals we should be aware that the minimization process will not trap into local minima of force field. It is not possible to avoid this problem completely, but we can minimize the probability to pass the global minima. Firstly, our minimizing algorithm should be appropriate for such kind of minimizations. It must be powerful enough to cross small local minimas and at the same time reasonably sensitive to fall into narrow, but deep minimas. We tried several minimization algorithms^{44,45}. Finally, we chose the Powell-Davidson algorithm^{46,47}. Parameters of optimization routine were optimized in each case and we saw that the process was very sensitive even to minor details of minimization algorithm. For example, parameters depended on a version of Fortran compiler and on an operating system using the same Fortran source code. The considerations in selecting parameters of minimizations are as follows: during the refinement procedure the equilibrium of structure, caused by different forces of force field, must be found. It is clear that this minimum of potential energy is not the equilibrium of different forces. On the one hand, the gradients of forces are remarkable, on the other, the minimization should not pass any significant minima.

A first step in estimating initial models for computation is the molecular modelling. Using physical models and computer graphics software we presumed

all kinds of possible conformations, according to unit cell measures. The chief idea was to avoid bad contacts. Based on this modelling system, the computer generated hundreds of initial models. These models were minimized by using the rigid-ring method. Results of these were checked on the basis of the local minimas reached and also by computer graphics facilities. Correspondingly, to check the results, a new set of initial models were generated and minimized. The results were subjected to yet another graphic check. On the basis of these results some minimization levels were sometimes added. The models obtained have already been subject to analysis and conclusions. In some cases an improvement of rigid-ring calculations MM3 calculations followed.

3.2.2. Parallel models of cellulose I crystalline structure

Several years ago VanderHart and Atalla⁴⁸ discovered that all native cellulose I-s are composed of two phases of crystals^{49,50}. Later, Sugiyama *et al.*⁵ described these phases in *Microdictyon tenuis*. They used electron diffraction technique and solved two unit cells: a triclinic one-chain unit cell for phase of Ia with cell parameters $a = 6.74 \text{ \AA}$, $b = 5.93 \text{ \AA}$, $c = 10.36 \text{ \AA}$, $\alpha = 117^\circ$, $\beta = 113^\circ$, $\gamma = 81^\circ$, and a monoclinic two-chain unit cell for phase of Ib with cell measures $a = 8.01 \text{ \AA}$, $b = 8.17 \text{ \AA}$, $c = 10.36 \text{ \AA}$, $\gamma = 97.3^\circ$. We refined the structures of these phases by using the rigid-ring method. A complete description of that work and its results are given elsewhere^{IV,V}. The structures have similar hydrogen bondings. It is possible to convert the structure of Ia into Ib via simple shifting of the chain sheets. The energy barrier calculated by the rigid-ring method is about 9 kcal/mol. This has an extremely high value. Calculations with annealing conditions would probably give less value. This high value can explain why the phase Ia does not transform into the Ib phase in normal conditions. Minimized energies are in agreement with

experimental data⁵¹. As cellulose Ia converts into the phase Ib during annealing process and cellulose I reforms into the phase II with mercerization process, the energy differences of 0.4 kcal/mol between Ia and Ib and 1.5 kcal/mol between Ib and II are extremely low. The difference between the phases of native cellulose are at the significant level. The hydrogen bond system is the same in native phases, although some investigators have reported changes in the system. From the point of view of energy calculations and modelling, it is hard to find another hydrogen bond system which would enable us to make comparisons with systems found by minimization (see Table IV.1). This statement is valid only for parallel structures of native celluloses derived from the data reported by Sugiyama *et al.*

3.2.3. Antiparallel models of native celluloses

Recent data of electron diffraction of native celluloses obtained by Sugiyama *et al* are explained as parallel structures. All chains in a unit cell are parallel. In one-chain unit cells there are no other possibilities, however, in the two-chain one, there exists an antiparallel option, too. The issue of parallelity of cellulose chains has been under discussion for several

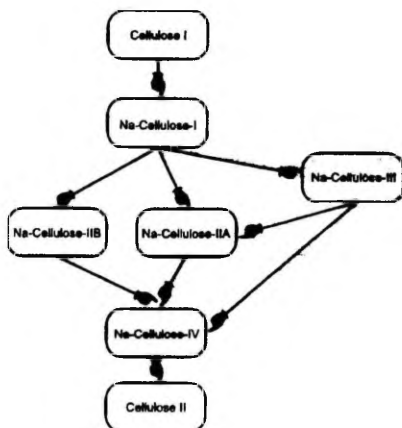


Figure 6. A conversion of native cellulose into cellulose II during a mercerization process.

decades^{52,53}. A number of researchers have given their interpretation, but no satisfactory explanation has been reached yet. According to the diffraction data^{36,54} and energy calculations^{55,56} it seems that cellulose II has an antiparallel structure. However, cellulose I was interpreted as parallel structure. During mercerization treatment native cellulose converts into cellulose II forming several intermediate structures. These phases have been thoroughly described by Sarko's group⁵⁷. They declare that already the first intermediate phase, the so-called Na-cellulose-I (see Figure 6), already has an antiparallel structure. It is extremely difficult to explain how it is possible to change the direction of a cellulose chain which has a molecular weight over 10000. Different researchers have elucidated this as interdigitation. There are different microcrystals in a cellulose fibre. Some of them are oriented *up*, some of them *down*. During the first step of mercerization process these crystallites will mix and form antiparallel structures. This process is complicated and has not been satisfactorily explained.

Another way to explain antiparallel cellulose II structures is to presume that cellulose I or at least some phases or components are already by themselves antiparallel. We have also proposed several models for the antiparallel structures of native celluloses. The construction of antiparallel models of unit cells is based on the parameters of unit cells as reported by Sugiyama *et al.* We constructed eight-chain unit cells⁵⁸ for both phases of native cellulose and simple antiparallel two-chain unit cell for the phase of $I\beta$. We calculated these models by using the rigid-ring method, improving several initial models by modelling and calculation methods^{VI,VIII}. These initial models are thoroughly described in^{VI,III}. Results of these calculations are presented in table VI.1.

An antiparallel two-chain unit cell has a relatively high energy level. At the same time, an eight-chain unit cell for the $I\beta$ phase, denoted as A3a (see Figure VI.10), has a low energy of -21.0 kcal/mol. However, antiparallel eight-chain unit

cell models for the $I\alpha$ phase do not offer good energy results. There exists an explanation why the $I\alpha$ phase needs not to be in global minima. A $I\alpha$ phase, existing independently from the $I\beta$ phase, has not been discovered yet. There appear only mixtures of two native cellulose phases. This explains why the $I\alpha$ phase can be found in structures which have comparatively higher crystal energies, for example, model A1a (see Figure VI.4). The model has a different hydrogen bond system from both the model P2 (see Figure VI.8) and the A3a (see Figure VI.10). This is in accordance with the hydrogen bond change in $I\alpha$ to $I\beta$ conversion marked by VanderHart and Atalla.

3.3. The full molecular mechanics (MM3) calculations of celluloses

3.3.1. Experimental

The starting coordinates of atoms were calculated by the rigid-ring method or taken from literary sources. Cellulose $I\alpha$ was described in^V, cellulose $I\beta$ in^V and crystalline small molecules as α -D and β -D-glucopyranose in⁵⁹. Minicrystals were constructed for calculating energies (see p.2.2 and in ^I). The molecular graphics software and some conversion software were used to construct crystals, to use the PLMR output data, to manipulate the molecules and to prepare MM3 input files. After minimization the initial and final structures were fitted by the least squares procedure. After the fitting procedure the average of the absolute values of differences between the initial and the final coordinates was reported as the *mean atomic movement*. This was also used for comparing different results of PLMR calculations.

3.3.2. Effect of the dielectric constant

As the MM3 routine use electrostatic interaction to model long term interactions, the value of the dielectric constant used will play an important role in

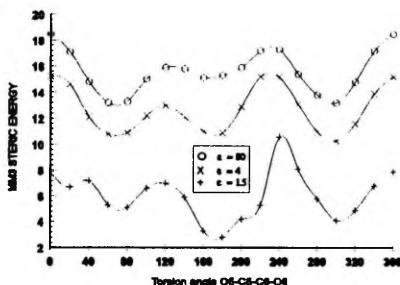


Figure 7. O6 rotation for β glucose. O6H orientation is left to optimize.

structure refinement. For that purpose we modelled several smaller molecules as glucose rings with different ϵ values. The results of varying the

dielectric constant in 'final steric energy' calculations are shown in Figure 1.5. The mean atomic movement is plotted in Figure 1.6. It is evident that the dielectric value close to 4 is the best solution. The latter gives the best agreement with experimental data²⁶. Dielectric values as low as 1.5 produce extreme movement, it even gives the best agreement with 6/31g* *ab initio* studies when ϵ is equal to 1. Figure 7 shows the simplest type of analysis, one for the rotation of primary alcohol group wherein the O6H orientation is left to optimize. We can see that the barrier does not depend much on the value of the dielectric constant.

3.3.3. Energies of cellulose polymorphs

The starting structures in MM3 calculations were taken from the best models of the PLMR process. The description of minicrystals built for minimization is given in¹. The model crystals were optimized with no restrictions of space group. The full molecular mechanics minimization process should be more valid than rigid-residue calculations. There are no restrictions to moving and flexing. The energy values calculated by MM3 are likely to be accurate; the standard deviation in calculated heat formation for 40 isolated alcohols and ethers was 0.38 kcal/mol. Energies of the minicrystals are probably less accurate. The results of MM3 minimization are given in Table V.3 and Table IV.1. The analysis of these results is described elsewhere^v. We got the energy of 185 kcal for *I α* phase and 182 kcal for *I β* phase. Analogous values of PLMR were -19.5 kcal/mol and -19.9 kcal/mol. For cellulose II the energy values were 176 kcal by MM3⁶⁰ and -21.4 kcal/mol by PLMR. These values are in good agreement with the experiment. The energies of both phases of cellulose I are slightly higher than the energy of cellulose II, and the *I α* phase is a bit higher than *I β* . The energies of both phases of native cellulose are extremely close. This may explain why they can coexist. The energy values of cellulose III and IV are comparatively higher. To compare the results of PLMR and MM3 we calculated the mean atom movement. These results are presented in Table V.3. Relatively low values show good agreement of both methods. Most of the best models of PLMR remain at their positions also after MM3 minimization. MM3 method found some additional intersheet hydrogen bonds in models of U₃, U₄ and U₆. PLMR did not find these bonds. Actually these models trapped into local minimas of a constrained force field. We also calculated lattice energies using MM3. For this purpose we removed the central chain and calculated energies of the chain and they remained 6 chains without minimization. These results are in Table

8. In comparison with other polymorphs, a lattice energy of cellulose IV is low. This means that intramolecular energy is high - the molecule has distortion from

equilibrium. The lattice energies, calculated by MM3, are very close to the values of PLMR, as most of the PLMR energy value component is an interchain energy. These MM3 values are slightly higher than PLMR interchain components. The differences between the full molecular mechanic and rigid-ring calculations do not render remarkably different results for the cellulose crystal

Table 8. Energies of cellulose models.

| Structures | Total energy (kcal/mol) | Lattice energy (kcal/tetrose) (/glucose) |
|----------------------------|----------------------------|---|
| Cellulose Ia | 185 | -76.1/(-19.0) |
| Cellulose Ib | 182 | -78.0/(-19.5) |
| Cellulose II | 176 | -79.0/(-19.8) |
| Cellulose III ₁ | 200 | -73.4/(-18.3) |
| Cellulose IV ₁ | 202 | -79.6/(-19.9) |
| Cellulose IV ₂ | 232 | -79.5/(-19.9) |

structure. Differences are minor and express in reality, the differences in force fields. Some differences may also occur as a result of the periodic boundary conditions used in rigid-ring calculations. These express more precisely the situation in a crystal. As we can see from the MM3 output, the side chains are in somewhat different conformations than the a central chain (see Figures II.1 and IV.2), the latter being in the most pseudoperiodic force field conditions.

3.4. Discussion over cellulose structure

It is clear that the issue of the structure of different cellulose polymorphs is far from being solved. The parallelity of cellulose chains and the structure of native celluloses remains the most dubious questions. The latter question has been given some light on, but serious problems still exist in the field.

It seems that the structure of cellulose II (*mercerized*) has an antiparallel structure. This is confirmed by different experiments and calculations. At the same time recent experiments reveal that cellulose I has a parallel structure. X-ray diffraction investigations of different intermediate states of mercerization process report that already the Na-cellulose-I has an antiparallel structure. It is explained as an interdigitation of chains, but this is not a very good interpretation. In our opinion the main problem in solving the structure of celluloses is to find the construction of unit cells and to solve the problem of parallelity of cellulose chains. An idea has been put forward that it is likely that native celluloses from different sources have unit cells with somewhat different parameter value. As our calculations show, very exact parameters of unit cells are not obvious. The results of the $I\beta$ unit cell refinement with parameters reported in *Pertsin et al.*⁶¹ and the results of the refinement of ramie celluloses^{60,62} are extremely close. It is more important to find a symmetry group and positions of cellulose chains. The problem of how to explain the conversion of cellulose I into cellulose II remains to be a common topic. An interdigitation is not a very satisfactory explanation. Atalla proposes that there can be two different cellulose II unit cells. This is one way to explain the situation. The other way to explain it is to look for antiparallel native cellulose structures. For these purposes we calculated several antiparallel structures based on unit cell measures reported in *Sugiyama et al.*⁶. As we found an antiparallel structure with

a very good energy, this explanation cannot be overlooked.

4. Conclusions

4.1. Methods

Analyzing the material presented in this paper, we can conclude that the structure of cellulose crystallites is mostly determined by hydrogen bonds and non-bonded interactions. Therefore, to refine these structures by using the steric energy minimization technique we can discard some of the components of full steric energy. The glucose ring is rigid enough to assume that it would remain fixed during a cellulose crystal structure refinement. Due to these simplifications, we can investigate the structural properties of cellulose chains in a crystal structure more precisely and faster.

Advantages of the rigid-ring method:

By decreasing the number of variables, we make the minimizing routine more efficient, i. e. the process will converge better.

By avoiding comparatively strong interactions in potential energy function, we make refinement routine more sensitive against weaker interaction which plays the main role in the formation of cellulose crystals. In fact the minimization process will not oscillate around strong interactions.

Drawbacks of the rigid-ring method:

The force field is deformed and does not correspond to the "real" force field.

Values of crystal energies that we get as a result have not enough physical background. It corresponds only to some terms of force field and have no accordance with experimental data.

Due to distortions in the force field there are more conformational barriers and local minimums compared to full molecular mechanics which makes the refinement of a global minimum more complicated.

We must be certain that the minimization would not remarkably affect the restricted degrees of freedom, and that the refining structure is not very sensitive to minor changes of these terms from equilibrium. It means that molecules in a crystal structure are not far from their conformation of "free" equilibrium.

As calculations improve, these simplifications will justify themselves. In the course of the calculations we got the right initial structure of cellulose I β , which was not achieved by the previous calculation with MM3 mostly due to the lack of complicated minimization functions. Even extremely simple potential energy functions give precise enough results.

The rigid-ring method seems to be suitable for fast and powerful research on crystal structures of celluloses and other polysaccharides. Even if the force field is deformed, it will not affect the results. Although this force field has more erroneous local minimas, the best results are easy to refine. It is possible to incorporate X-ray diffraction data for this refinement process, though, it seems that the potential energy calculations are very powerful even without the diffraction data.

The construction of minicrystal models for modelling cellulose crystals is valuable, even if it has drawbacks which were mentioned above. Incorporating X-ray data into this minimization technique makes it more awkward. At the same time calculations with full molecular mechanics give us a possibility to compare our results with experimental data, i.e. comparison of heat of formation. Similarly, a full molecular mechanics studies may give more detailed information about the structure. For final refinement purposes it would be more useful as it gives more accurate results. Larger crystal models would be better but they need extremely

large computer resources. However, for structure refinement purposes it suffices with the rigid-ring calculations.

4.2. Structure of celluloses

As we have mentioned several times, the complete solving of cellulose polymorphs structure is not yet finished. Cellulose II seems to have an antiparallel structure. The crystal energy calculations of both phases of native cellulose give results which are in a good agreement with the experimental data on the conversion $I\alpha \rightarrow I\beta \rightarrow II$. Though the results of the parallel structure of native celluloses is very easy to interpret, it is hard to explain the conversion of a parallel structure into an antiparallel structure during mercerization process. According to this we cannot overlook the idea about antiparallel structures of native celluloses. An alternative way is to reinspect the conception of cellulose II. Only further experiments will be able to answer that problem.

* * *

In the present work the rigid-ring methodology has been developed to refine structures of celluloses and other polysaccharides. This routine has been improved by several ways and compared with full molecular mechanics calculations. Several calculations have been made by using full molecular mechanics. Also, structures of both polymorphs of celluloses refined by using both the rigid-ring method and the full molecular mechanics refinement. Several hypothesis on the structure of polymorphs of native celluloses were set up according to the results of molecular modelling and rigid-ring calculations.

Acknowledgements

I am very thankful to Alfred D. French for the productive collaboration. Also, I am grateful to Raik-Hiio Mikelsaar and Alexander J. Pertsin for their excellent ideas and for supervising the present research. I thank Aksel Haav for helping me during my first steps in crystallography. Also, I wish thank Normann L. Allinger for providing me with the MM3 program.

References

1. Sarko, A.; Southwick, J.; Hayashi, J. *Macromolecules* 1976, **9**, 857.
2. Chanzy, H.; Imada, K.; Vuong *Protoplasma* 1978, **94**, 299.
3. Smith, P. J. C.; Arnott, S. *Acta Crystallographica* 1978, **A34**, 3
4. Millane, R. P.; Narasaiah, T. V. *Polymer* 1989, **60**, 1763.
5. Sugiyama, J.; Vuong, R.; Chanzy, H. *Macromolecules* 1991, **24**, 4108.
6. Stipanovic, A. J.; Sarko, A. *Macromolecules* 1976, **9**, 851.
7. Sugiyama, J.; Okano, T.; Yamamoto, H.; Horii, F. *Macromolecules* 1990, **23**, 3196.
8. Yamamoto, H.; Horii, F.; Odani, H. *Macromolecules* 1989, **22**, 4130.
9. Marrinan, H. J.; Mann, J. *J Polymer Sci* 1956, **21**, 301.
10. Horii, F.; Hirai, A.; Kitamaru, R. *ACS Symp Series* 1984, 260, 27.
11. Bock, K. *Pure and Applied Chemistry* 1983, **55**, 605.

12. Homans, S. W.; Dwek, R. A.; Rademacher, T. W. *Biochemistry* 1987, **11**, 1465.
13. Cherniak, R.; Jones, R. G.; Reiss, E. *Carbohydrate Research* 1988, **172**, 113.
14. Golebiewski, A.; Parczweski, A. *Chem Rev* 1974, **74**, 519.
15. Tvaroška, I. *Pure and Applied Chemistry* 1989, **61**, 1201.
16. Pullman, B. *Quantum Mechanics of Molecular Conformations*, 1976, Wiley, New York.
17. Tvaroška, I.; Bleha, T. *Chem Papers* 1985, **39**, 805.
18. VanGunsteren, W. F.; Weiner, P. K. *Computer Simulation of Biomolecular Systems*, Leiden, 1989.
19. Hardy, B. J.; Sarko, A. in *Cellulosics: Chemical, Biochemical and Material Aspects*, Eds. Kennedy, J. F.; Phillips, G. O.; Williams, P. A.; Ellis Horwood Ser. Polymer Science and Technology, Ellis Horwood, New York, 1993, 41.
20. Mikelsaar, R.-H. *Trends in Biotechnology* 1986, **4**, 162.
21. Boyd, D. B.; Lipkowitz, K. B. *J Chem Educ* 1982, **59**, 269.
22. Burkert, U.; Allinger, N. L. *Molecular Mechanics*, ACS Monograph 177, ACS, Washington D. C., 1982.
23. Ha, S. N.; Giammona, A.; Field, M.; Brady, J. W. *Carbohydrate Research* 1988, **180**, 207.
24. Weiner, S. J.; Kollman, P.A.; Case, D. A.; Shandra Singh, U.; Ghio, C.; Alagona, G.; Prefeta, S. Jr.; Weiner, P. *J Amer Chem Soc* 1984, **106**, 765.
25. Allinger, N. L.; Rahman, M.; Lii, J.-H. *J Amer Chem Soc* 1990, **112**, 8293.
26. Allinger, N. L.; *J Amer Chem Soc* 1977, **99**, 8127.
27. Lii, J. H.; Allinger, N. L. *J Comput Chem* 1991, **12**, 186.
28. Lii, J. H.; Allinger, N. L. *J Amer Chem Soc* 1989, **111**, 8576.
29. Allinger, N. L.; Yuh, Y. H.; Lii, J.-H. *J Amer Chem Soc* 1989, **111**, 8551.
30. Lii, J. H.; Allinger, N. L. *J Amer Chem Soc* 1989, **111**, 8566.

31. Arnott, S.; Scott, W. E. *J Chem Soc Perkin II* 1972,324.
32. French, A. D.; Dowd, D. in *Ceilulosics: Chemical, Biochemical and Material Aspects*, Eds. Kennedy, J. F.; Phillips, G. O.; Williams, P. A. Ellis Horwood Series Polymer Science and Technology, Ellis Horwood, New York 1993, 51.
33. Zugermaier, P.; Sarko, A. in *Fiber Diffraction Methods*, Eds. French, A. D; Gardner, K. H.; ACS Symposium Series 141, ACS Books, Washington DC 1980, 225.
34. Pertsin, A. J.; Kitaigorodsky, A. I. *The Atom-Atom Potential Method Applications to Organic Molecular Solids*, Springer Ser. Chem. Phys., V43, Springer-Verlag, Berlin, 1987.
35. Kolpak, J. F.; Weih, M.; Blackwell, J. *Polymer* 1978, 19, 123.
36. Дашевский, В.Г. *Конформационный анализ макромолекул*. Москва: Наука 1987.(in Russian).
37. Дашевский, В. Г. *Конформационный анализ органических молекул*. Москва: Химия. 1982.(in Russian).
38. Mirskaya, K. V.; Kozlova, I. E.; Bereznitskaya, V. F. *Phys Stat Solid* 1974, 62, 291.
39. Kitaigorodsky, A. I.; Mirskaya, K. V.; Nauchitel', V. V. *Kristallografia* 1969, 62, 291.
40. Hamilton, W. C. *Acta Crystallographica* 1965, 18, 502.
41. Pertsin, A. J.; Nugmanov, O. K.; Marchenko, G. N.; Kitaigorodsky, A. I.; *Polymer* 1984, 25, 107.
42. French, A. D.; Murphy, V. G. *Carbohydrate Research* 1973, 27, 391.
43. Chu, S. S. C.; Jeffrey, G. A. *Acta Crystallographica* 1970, B26, 1373.
44. Березин, И. С.; Жидков, Н. П. *Методы вычисления*. Москва 1959.(in Russian).
45. Himmelbau, D. M. *Applied Nonlinear Programming*, McGraw Hill, 1979.
46. Powell, M. J. D. *Comput J* 1964, 7, 155.
47. Zangwill, W. I. *Comput J* 1967, 10, 293.

48. VanderHart, D. L.; Atalla, R. H. *Macromolecules* 1984, **17**, 1465.
49. Atalla, R., H.; VanderHart, D. L. *Science* 1984, **223**, 283.
50. Horii, F.; Hirai, A.; Kitamaru, R. *Macromolecules* 1987, **20**, 2117.
51. Ranby, B. G. *Acta Chem Scand* 1952, **6**, 101.
52. Gardner, K. H.; Blackwell, J. *Biopolymers* 1974, **13**, 1975.
53. French, A. D. in *Polymers for Fibers and Elastomers*, Ed. by Jett C. Arthur, Jr., ACS Symp Series V26, 43.
54. Kolpak, J. F.; Blackwell, J. *Macromolecules* 1976, **9**, 273.
55. Nishimura, H.; Okano, T.; Sarko, A. *Macromolecules* 1991, **24**, 759.
56. Nishimura, H.; Sarko, A. *Macromolecules* 1991, **24**, 771.
57. Sarko, A.; Nishimura, H.; Okano, T. in *The Structure of Cellulose*, ACS Symp Series **340**, ACS Washington DC, 1987, 169.
58. Honjo, G.; Watanabe, M. *Nature* 1958, **181**, 326.
59. Brown, G. M.; Levy, H. A. *Acta Crystallographica* 1979, **B35**, 656.
60. Woodcock, C.; Sarko, A. *Macromolecules* 1980, **13**, 1183.
61. Pertsin, A. J.; Nugmanov, O. K.; Marchenko, G. N. *Polymer* 1986, **27**, 597.
62. French, A. D.; Howley, P. S. in *Cellulose and Wood - Chemistry and Technology*, Ed. Schuerch, C. John Wiley & Sons, New York 1989, 159.

List of publications

- I. French, A. D.; Dowd, D.; Aabloo, A. *Int J Biol Macromolecules* 1993, **15**, 30-36.
- II. Aabloo, A.; Pertsin, A. J.; Mikelsaar, R.-H. in *Cellulosics: Chemical Biochemical and Material Aspects*, Eds. J. F. Kennedy, G. O. Phillips, P. A. Williams, Ellis Horwood Series Polymer Science and Technology, Ellis Horwood, New York, 1993, 61-65.
- III. Mikelsaar, R.-H.; Aabloo, A. in *Cellulosics: Chemical, Biochemical and Material Aspects*, Eds. J. F. Kennedy, G. O. Phillips, P. A. Williams, Ellis Horwood Series Polymer Science and Technology, Ellis Horwood, New York, 1993, 57-60.
- IV. Aabloo, A.; French A. D. *Makromolek Chem, Theory and Simulations* 1994, **2**, in press.
- V. Aabloo, A.; French A. D.; Mikelsaar, R.-H.; Pertsin, J. *Cellulose*, in press.
- VI. Mikelsaar, R.-H.; Aabloo, A. *Cellulose*, in print.
- VII. Aabloo, A.; French, A. D.; Mikelsaar, R.-H. in *Cellulose Derivates and Related Polysaccharides*, Eds J. F. Kennedy et al, Ellis Horwood Series Polymer Science and Technology, in press.
- VIII. Mikelsaar, R.-H.; Aabloo, A. in *Cellulose Derivates and Related Polysaccharides*, Eds. J. F. Kennedy et al, Ellis Horwood Series in Polymer Science and Technology, in press.

Tselluloosi kristalsete faaside struktuuri uurimine kasutades energeetilisi arvutusi

Tselluloos on üks levinumaid biopolümeere maailmas. Tema struktuuri on uuritud juba aastakümneid, kuid sellest hoolimata leidub veel palju lahendamata probleeme. Kristallne tselluloos esineb erinevates vormides - polümorfides. Looduslik tselluloos (I) koosneb kahest komponendist nn. faasid α ja β . Sõltuvalt päritolust sisaldab looduslik tselluloos neid komponente erinevas vahekorras. Kõige levinum tööstuslik tselluloosi vorm on tselluloos II. Teised faasid on vähem levinud.

Tselluloos I kristalsete faaside ühikrakkude ehitus on määratud hiljuti, seetõttu ongi käesoleva töö üheks eesmärgiks nende faaside struktuuri uurimine. Samuti pakume välja mõningaid alternatiivseid ühikraku struktuure, mis seletaksid paremini tselluloosi faaside üleminekuid. Struktuuri määramiseks kasutame kristalli potentsiaalse (*steerilise*) energia arvutusi. Arvutuste käigus me fikseerisime need konformatsioonilised parameetrid, mis minimiseerimise käigus niikuinii oluliselt ei varieeru ning seetõttu ainult segavad arvutusi. Samuti kasutasime kristalli konstrueerimise käigus perioodilisi ääritingimusi jms., mis vastab paremini reaalsele jõuväljale, mille paikneb suhteliselt pikk polümeerahel. Arvutusmetoodikat kontrollisime ka molekulaarmehaanika meetodiga. Tulemused langevad vägagi hästi kokku.

Miniature crystal models of cellulose polymorphs and other carbohydrates

Alfred D. French*

Southern Regional Research Center, PO Box 19687, New Orleans, Louisiana 70179, USA

Donald P. Miller

Consultant, PO Box 423, Waveland, Mississippi 39576, USA

and Alvo Aabloo

Department of Experimental Physics, Tartu University, 202400 Tartu, Estonia

(Received 10 September 1992; revised 28 September 1992)

Miniature crystal models of cellulose and other carbohydrates were evaluated with the molecular mechanics program MM3. The models consisted of groups of 24 to 32 monosaccharide residues, with the models of mono- and disaccharides based on well-established, single-crystal work. Structures of the cellulose forms and cellotetraose were based on published work using fibre diffraction methods. A structure for the single-chain Ia cellulose unit cell was also tested. A dielectric constant of about 4 was best for this type of work. Calculated intra- and intermolecular energy for glucose agreed with literature values for the heat of combustion. Cellulose II had the lowest calculated energy for a cellulose form, followed by Ia, cellulose III₁, ramie I, IV₁₁ and IV₁. Optimization of cellulose IV caused larger mean atomic movements from the original crystallographic positions than the other cellulose forms, and cellotetraose had larger movements than any of the other structures. Lattice energies for the cellulose forms were about 20 kcal/mol of glucose residues, with a dominant van der Waals component.

Keywords: Cellulose; crystal models; molecular mechanics

Introduction

Pure cellulose crystallizes in various forms, named I to IV, depending on the history of the sample. Cellulose I, the major native type, has recently been recognized to occur mostly as mixtures of the two subclasses, Ia and Ib¹. The two subclasses occur in different amounts and have, respectively, one and two chains per unit cell. Together² they account for the 8-chain unit cell proposed earlier³. Cellulose II results from mercerization (treatment in 22% sodium hydroxide) or crystallization from solution. Cellulose III, the product of treatment of cellulose I or II with liquid ammonia or other amines, has two subclasses, III₁ and III₂, depending on the parent structure. Finally, cellulose IV results from treatment at high temperature (in glycerol at 260°C) of I, II or III, with subclasses IV₁ and IV₁₁ depending on the parent structure. Cellulose IV also appears in immature native samples⁴.

Researchers have long thought that these various forms have different relative stabilities, with the most stable form being cellulose II. Of the native Ia and Ib forms, the product of annealing a mixture is pure Ib, so Ib is thought to have the next lowest energy⁵. Since the III and IV forms can revert to their parent I or II forms, they are thought to have slightly higher energy than the parent forms.

As far as we are aware, quantitative comparison of the

energies of the various forms has not been attempted before. Instead, these differences in stability are normally ascribed to different intra- and intermolecular hydrogen bonding schemes, often proposed from X-ray fibre diffraction experiments. However, such experiments are so difficult that even the chain-packing polarity is often not well-determined. Even in the far more accurate X-ray diffraction studies of single crystals of small molecules, the hydrogen bonding is often difficult to assess. If accurate positions of hydrogen atoms are needed, low temperatures and/or neutron diffraction are used. Thus, hydrogen bonding schemes resulting from fibre diffraction studies are speculative. In fibre diffraction work, hydrogen bonding is derived with the aid of computer models. These models are necessary components of the fibre diffraction method, but typically come from software that is less well developed than other software used only for modelling.

In the present work, we have studied the energies of small model crystals of most of the cellulose forms, using a sophisticated molecular mechanics system, MM3^{3,6}, that was designed to handle a wide variety of organic molecules. For comparison, we have modelled several crystal structures of smaller carbohydrate molecules for which good crystal structure data are available. A variety of information comes from these studies. An approximate lattice energy (or heat of sublimation) can be determined from the optimized microcrystal by first removing the central molecule. The energy of the central molecule and the total energy of the surrounding molecules are then calculated (without further optimization). Their sum exceeds the energy of

*To whom correspondence should be addressed.

This work is the property of the US Government and is not subject to copyright

micrystal by the amount of the lattice energy. Lattice energy can be added to the heat of formation of isolated molecule (calculated as a normal option (3) giving a total ΔH_f , which can be compared with the values for the heat of formation of solid hydrates.

Lattice energy can be broken down into the van der Waals and 'dipole-dipole' terms. These correspond to the dispersive forces and the electrostatic forces. From the solubility of cellulosic molecules in many solvents, it is clear that both are relatively strong, but we are not aware of specific proposals of their relative values.

The total steric energy reported by MM3 is a sum of the various attractive and repulsive terms. It includes the bond stretching and angle bending costs of forming pyranose and furanose rings, as well as non-bonded forces that can apply to both intra- and intermolecular interactions. The more stable conformation will have a lower total energy, regardless of whether the stability comes from a more stable isolated molecule or from a better intermolecular arrangement.

To assess the stabilities of the various cellulose forms, we have calculated the total steric energy terms for each conformation. If the energy is much higher than for another, similarly sized group of molecules, then the proposed structure may not be valid.

For the micrystals are optimized, there will have been some changes in atomic positions from the original coordinates, even when starting with well-determined structures. Lii and Allinger, who used a special program (CRSTL) based on the non-bonded terms for modelling crystals⁷, found adjustments of bond lengths of as much as 0.28 Å for relatively small hydrocarbon molecules. Those differences were attributed to limiting assumptions such as a spherical potential for atoms. While some movement is therefore expected, excessive movement would indicate a defective model structure.

The choice of dielectric constant is critical in MM3 calculations, wherein it scales the electrostatic interactions relative to the other forces. In MM3(90), the dipole-dipole energy, used instead of energy schemes based on explicit atomic charges in other modelling programs, depends on the dielectric constant, as does a lone pair hydrogen bonding term that is reported as part of the dipole-dipole value. A value near 4 was suggested

when MM3 results were compared with crystal structures, to mimic the effect of crystalline environments on isolated molecules⁸. However, we were explicitly creating a miniature crystal, and the question therefore arose whether the optimum balance between the dispersive and electrostatic forces would be obtained with 4 or with some other value.

Experimental

Starting coordinates

We started with the proposed atomic coordinates of the crystal structures already in the literature, or tables deposited with the structure reports. For the cellulose structures, we chose for internal consistency mostly the work from Sarko's group at Syracuse⁹⁻¹². Cellulose Ia was described elsewhere¹³. Crystalline small molecules included α -D-¹⁴ and β -D-glucopyranose¹⁵, β -D-fructopyranose¹⁶, methyl β -D-galactopyranoside¹⁷, sucrose¹⁸ and β -D-cellobiose¹⁵. We also modelled a crystal of cellotetraose proposed from fibre diffraction studies¹⁹. Some results of a similar miniature crystal study were reported elsewhere for the tetrasaccharide nystose²⁰.

Construction of mini-crystals

Our models of the cellulose forms consisted of seven cellotetraose molecules arranged by the applicable symmetry operators to make a pseudohexagonal, close-packed miniature crystal (Figure 1). The small molecules were similarly arranged (see Figure 2), depending on the crystal structure, so that a central molecule was surrounded, in so far as possible, on all sides. Terminal hydroxyl hydrogen atoms or hydroxyl groups were added to the cellulose models as needed, with the orientations of the new OH groups being essentially random. Other OH groups of the cellulose models were oriented to build a network of hydrogen bonding corresponding to the authors' proposals. In the case of cellotetraose, absent a proposed hydrogen bonding scheme, a network was devised. These micrystals were then optimized, without restrictions of any kind on the atomic movement.

Calculations

The calculations were done with IBM-PC compatible 486 and VAX computers. The 1990 version of MM3 was

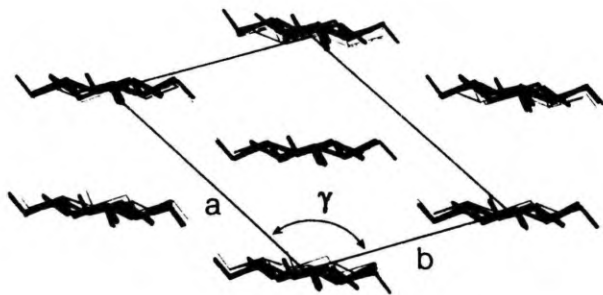


Figure 1 The micrystal model of cellulose III_c, before (---) and after (—) optimization with MM3

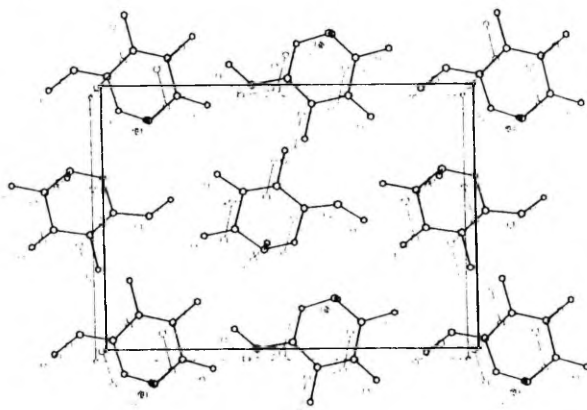


Figure 2 The starting minicrystal of α -D-glucopyranose with 27 molecules

used, with the energy-based (rather than geometry-based) default termination criterion that depends on the number of atoms in the structure (0.05 kcal/mol for a 27-molecule minicrystal). The CHEM-X program²¹ was used to construct the crystals, to manipulate the molecules, and to prepare MM3 input files. CHEM-X was also used to fit the initial and final structures by a least squares procedure, minimizing the function

$$F = \sum (a_i - b_i)^2$$

where a_i and b_i are the atom coordinates of the i th atom after and before minimization²². After this fitting procedure, the average of the absolute values of the residual differences between a_i and b_i was reported as the mean atomic movement. It was computed only for the oxygen and carbon atoms.

Limitations

Several limitations characterize our models. The number of molecules is small, as is the chain length. The limits on size arise because of the limit of 700 atoms in the MM3 program. While that size limit is somewhat artificial, significantly larger models would require much more computer time. For example, one more layer of molecules around the current 27-molecule crystal of glucose would add 98 molecules. The larger model would require about 20 times as long to optimize.

Energy increases

In these and other modelling studies of carbohydrates, MM3 often has a problem with energy increases that terminate the minimization. This happens only when using the block diagonal least squares minimizer, necessary for structures containing more than 80 atoms. Figure 3 shows the energy values during three separate minimization runs of the glucose minicrystal derived from the neutral diffraction study. On the 71st cycle of optimization, the energy went up slightly. Optimizations of structures that yield energy increases also indicate

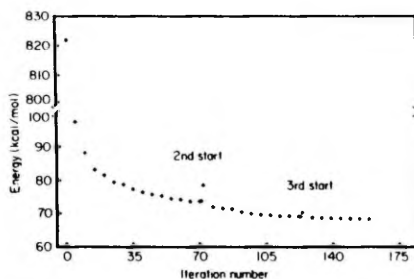


Figure 3 Energy minimization of α -D-glucose at a dielectric constant of 4. The restarts are indicated. After the energy went up at the end of the first run, the hydroxyl groups showing maximum movement were rotated and the optimization restarted. At the end of the second run, there was still substantial fluctuation of some other hydroxyl hydrogen atoms (see Figure 4). The energy declined only slightly during the third run

severe oscillations of hydroxyl hydrogen atoms on the output, as the atoms with maximum atomic movement, shown in Figure 4. The MM3 output lists, at each five iterations, which atom would be moved the most and the projected extent of the movement, based on the derivatives calculated during the optimization step. The atomic movement per iteration is actually limited to about 0.25 Å and does not occur to the extent shown in Figure 4. This problem prevented the full optimization of several of the minicrystals, at least until the indicated hydroxyl groups were rotated (with CHEM-X) to a suitable alternate staggered orientation. After rotation of the hydroxyl groups, the energy would be momentarily higher, such as the 5 kcal/mol increase shown for iteration 72 in Figure 3, but minimization would quickly

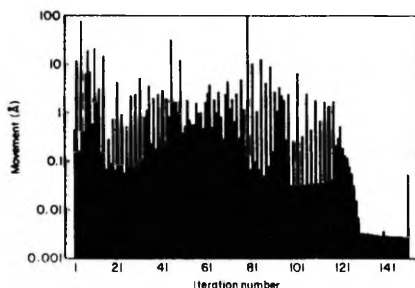


Figure 4 The maximum movement by any atom, as indicated by MM3. The movement reached a suitable value only during the third run, even though the energy changed only slightly. The atoms are limited to movements of about 0.25 Å by the program; they are not actually moved as much as indicated. As in Figure 3, the first run ended at 71 iterations, and the second at the 118th iteration

lower the total to a value less than the previous low. The problem happened often with the somewhat random orientation given to the added hydroxyl groups at the cellotetraose ends, but sometimes it also afflicted other groups on the surfaces of the minicrystals.

The maximum movement of any individual atom indicated in Figure 4 is distinct from the mean atomic movement that we have used to indicate the quality of the force field and, in the case of the structures determined by fibre diffraction, the quality of the structure. The mean atomic movement refers to the average of all the distances between the initial and final positions of each atom, and is only applied to the carbon and oxygen atoms in this work. The maximum atomic movement refers to the atom that is indicated to need to move the most to lower the energy during minimization. The atom that is indicated to have the greatest movement is almost always a hydroxyl oxygen when the problem with energy increase during minimization occurs.

When this problem causes premature termination of the optimization, the differences in the initial and final atomic positions are not as large as if the structure were fully optimized. Also, the energy is not as low as when the optimization has been carried full term after rotation of the indicated groups. Of course, part of the increased movement and decreased energy is due to the rotation of the hydroxyl hydrogen atoms. In a few instances, it was not possible to find an alternative position that allowed full optimization. However, after several trials, even the structures that continued to terminate with an energy increase were thought to be reasonably well optimized. Therefore, we reluctantly accepted these results as the best available.

Results and discussion

Effect of varied dielectric constant

Results of varying the dielectric constant in the MM3 calculation are shown in Figures 5 and 6. Figure 5 shows the variation in calculated energy. In view of its 880 kcal/mol range, the importance of the dielectric constant would be hard to ignore. For both glucose

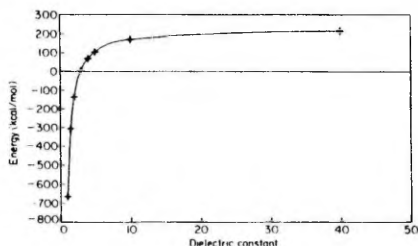


Figure 5 Plot of the MM3 'final steric energy' for the minicrystal models of α -D-glucose at different dielectric constants

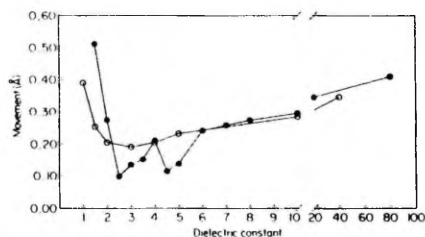


Figure 6 Plot of the mean atomic movement for the minicrystal models of glucose at different dielectric constants. (○) α -D-glucose; (●) β -D-glucose

molecules, the mean atomic movement during optimization seems to have a general minimum centred around a dielectric of 4, as shown in Figure 6. β -D-Glucopyranose has lower values on either side of 4, but they may be due to incomplete minimization of those structures. Therefore, a value of 4 was used on the other models in this study. Dielectric values as low as 1.5 produced extreme movement, even though MM3 agreed best with 6/31g* *ab initio* studies when the dielectric constant was set to 1.0²³. When condensed phase systems are modelled with molecular mechanics programs using a dielectric constant of 1, the resulting imbalance between the electrostatic forces and all other energy terms is a likely source of error. This observation also suggests that *ab initio* studies and condensed phase experiments will not have comparable results.

Movements

Table 1 shows the mean atomic movements that result from the optimization with MM3 of the different crystal structures. The structures based on good single crystal data have a range of 0.09 to 0.21 Å in the mean atomic movements. There is a slightly wider range for the cellulose structures and a large movement for cellotetraose. The two cellulose IV structures, in particular, have large mean movements. Inspection of the initial and final structures showed that the primary alcohol groups in both cellulose IV₁ and IV_{II} moved substantially during optimization, accounting for much of the mean

Table 1 Models, movements (Å) and energies (kcal/mol) for minicrystals

| Compounds modelled: | Model size | Movement | Total energy (kcal) | Lattice energy (kcal) /tetraose (/glucose) |
|-------------------------------------|-------------|----------|---------------------|--|
| Cellulose Ia | 7 tetramers | 0.106 | 185 | |
| Ramie cellulose I | 7 tetramers | 0.160 | 201 | -76.1/ -19.0 |
| Cellulose II | 7 tetramers | 0.202 | 176 | -79.0/ -19.75 |
| Cellulose III ₁ | 7 tetramers | 0.198 | 200 | -73.4/ -18.3 |
| Cellulose IV ₁ | 7 tetramers | 0.255 | 209 | -79.6/ -19.9 |
| Cellulose IV _{II} | 7 tetramers | 0.254 | 202 | -79.5/ -19.9 |
| Cellotetraose | 8 tetramers | 0.418 | 232 | |
| | | | | Energy/molecule |
| α -D-Glucose | 27 monomers | 0.205 | 68 | -37.4 |
| β -D-Glucose | 27 monomers | 0.211 | 36 | -37.6 |
| β -D-Fructopyranose | 25 monomers | 0.118 | 97 | -35.8 |
| Methyl β -D-galactopyranoside | 27 monomers | 0.091 | 114 | -36.1 |
| β -D-Cellobiose | 16 dimers | 0.092 | 199 | -61.0 |
| Sucrose | 14 dimers | 0.209 | 296 | -44.5 |

movement. The considerable atomic movement in the cellotetraose structure suggests that the proposed structure may not be correct, as indicated by the authors. The conformations of some of the model cellotetraose molecules changed to structures with approximate two-fold screw symmetry. Those changes in conformation, however, could have resulted instead from our guesses regarding the hydrogen bonding in the structure. In retrospect, it seems that having more specific guidance to the proposed hydrogen bonding would be worthwhile, even though the hydrogen bonding is not rigorously determined. A result consistent with the original authors' opinion would be useful for tests such as the present work.

The mean atomic movement in the model structures was not isotropic. Superimposition of the optimized model and initial structure showed that there was very little movement along the fibre axis. In the (100) planes of the ramie cellulose I, III (see Figure 1) and IV structures that contain the inter-chain hydrogen bonding, there was also little movement. The majority of movement was perpendicular to those planes. The (100) planes were slightly separated, as if the hydrophobic bonding were not pulling the structure together strongly enough. However, this may be a result of the small model size. The variations in unit cell dimensions reported for different native celluloses²⁴ may be due in part to the different crystallite sizes found in the different materials. Small crystallites would give slightly larger unit cells because of fewer long-range interactions pulling the structure together.

Movement in the cellulose II lattice was assessed by rotating the structures 74°, so that the planes of the chains were parallel to the x-axis. The components of movement were separated and, again, the major movement occurred perpendicular to the main planes of the cellulose chains. The movement along the z-axis (the fibre axis) was 0.058 Å, along the x-axis was 0.083 Å, and along the y-axis of Cartesian space it was 0.149 Å. Because the lengths in MM3 for the glycosidic C-1-O-1 bond are off by as much as 0.02 Å²⁵, the deficiencies in the z-direction, to which the glycosidic bond are nearly parallel, are small. Similar calculations for β -D-fructopyranose showed x-, y- and z-axis movements of 0.043, 0.054 and 0.074 Å,

respectively. In that crystal structure, the planes of the rings are perpendicular to the z-axis, along which the greatest movement is observed.

Energies

Table 1 also shows the total steric energies and heats of fusion (when calculated). The energy values for the different structures vary widely, even accounting for the differences in the sizes of the crystals needed to provide a totally surrounded central molecule. As a check on the overall validity of these calculations, we have also computed the heat of formation of β -D-glucose with MM3. The isolated β -D-glucose molecule in its lowest energy conformation has a ΔH_f of -265.8 kcal/mol, calculated at a dielectric constant of 1.5. When added to the -37.6 kcal/mol of fusion energy, the total of -303.4 equals the literature value of -303 kcal/mol calculated from the heat of combustion for solid glucose²⁶. The dielectric constant is somewhat less critical for the intramolecular energy. At a dielectric constant of 4, the isolated β -D-glucose molecule has a ΔH_f of -260.0 kcal/mol, still allowing a sum reasonably close to the literature value.

Energies of cellulose polymorphs. Of the cellulose structures, cellulose II has the lowest energy, resulting from both the low intramolecular energy, and as shown by the lattice energy, the second-lowest intermolecular energy. Its antiparallel model has four up (corner) chains and three (down) central chains. The corner chains have O6 in the *gt* position at slightly lower energy than the *tg* position occupied on the central chains. (The lower energy for the *gt* form is a consequence of the dielectric constant of 4, as well.) The energy of the packing model with four down chains and three up chains should also be averaged with this one. When this is done, the total energy increases about 5 kcal, still leaving cellulose II with the lowest energy. Only the ramie cellulose model differs from expectations, having energy substantially higher than the Ia model, equal to the III₁ model. The lattice energy of cellulose III₁ is broken down in Table 2 into the van der Waals and dipole-dipole terms for the

Table 2 Breakdown of energy calculations for cellulose III_I

| Energy term | 7-tetramer minicrystal | 6 outer chains | 1 inner chain | Lattice energy /tetraose (/glucose) |
|---------------|---------------------------|-------------------|------------------|--|
| van der Waals | -189.2 | -117.9 | -9.0 | -62.3 (-15.6) |
| Other | | | | |
| Dipole-dipole | -58.7 | -42.6 | -4.9 | -11.2 (-2.8) |
| Total | | | | -73.4 (-18.3) |

entire model, the isolated centre molecule and the remaining model. It shows that, computed at a dielectric constant of 4, the magnitude of the van der Waals attractions (15 kcal/mol of glucose residues) is considerably larger than the values for electrostatic attraction (3 kcal/mol of glucose residues). Thus, in the past, the contributions by the hydrophobic groups to crystal stability have been underestimated. Hydrogen bonds have both electrostatic and van der Waals components. If the single hydrogen bond in the cellulose III model has an energy of 5 kcal/mol, the remaining lattice energy would be about 13 kcal/mol, all arising from van der Waals forces.

The lattice energy of IV_I is slightly lower than all the other forms, while its overall energy is greater. This suggests that the intramolecular energy term is higher.

Mono- and oligosaccharide energies. The monosaccharide structures have substantially lower (more negative) lattice energies per residue than the larger models, owing to the larger number of intermolecular hydrogen bonds that are possible in monosaccharides. Cellobiose has less stability per residue than glucose, partly because two of the hydroxyl groups of glucose have been replaced by the glycosidic linkage in the disaccharide. The model sucrose disaccharide has a high energy compared with the cellobiose model. Since the sucrose model had only 14 disaccharide residues in a structure with P2₁ symmetry, the packing in its minicrystal was similar to that of the cellulose models, except that there were two layers. Therefore, some of the three-dimensional character of the small-molecule structures was lost. A better model of sucrose would have a third layer of molecules above or below our arrangement, but this was beyond the capacity of our method. The high energy can also be attributed to two other factors. One is the presence of furanose rings, which have significantly higher angle-bending and torsional energies than pyranose rings (amounting to several kcal/mol per residue). Another factor is the presence of overlapping anomeric effects at the sucrose linkage. Other work has shown that MM3 may overestimate the energies of such linkages, although the increase for sucrose is not as severe as for other crystals, such as raffinose, where the largest problem exists²⁷. In the models of nystose²⁰, where the sucrose linkage has an MM3 energy 3 kcal/mol higher than the global minimum, the mean atomic movement in the minicrystal was 0.312 Å, although the movement for the central molecule in the crystal was smaller. A partial reason for the smaller lattice energy in the sucrose model compared with may be that crystalline sucrose has two intramolecular hydrogen bonds, diminishing the opportunity for intermolecular hydrogen bonds.

Conclusions

The construction of model minicrystals and the calculation of their energies seems to be a valuable type of modelling study. We have shown that it is possible to obtain nearly quantitative values of the heat of formation, an important experimental quantity. Also, the approach provides a useful means to determine the value of the dielectric constant used to scale the electrostatic and other forces in a molecular mechanics force field. The work showed that a value of 4 is about right; the consequences of excessive atomic movement and decreased energy from using a value of 1 or 1.5 to model condensed phases are made quite clear in this study.

The minicrystal method also seemed to be useful for testing structures proposed in fibre diffraction studies. The proposed structures of cellulose IV had substantial movement of the primary alcohol groups and were thus the least compatible with the MM3 force field. We were also unable to supplement the proposed carbon and oxygen positions for cellotetraose with a hydrogen bonding scheme that avoided substantial movement during optimization with MM3. The authors had concluded that the cellotetraose data were not solved and refined to an acceptable degree of accuracy. Like any modelling study, the minicrystal technique cannot prove that a structure is correct, nor, because of the extensive time required for optimization, is it useful for finding the range of likely possibilities. Instead, it seems to function as a sophisticated secondary check of the modelling component of structures derived by fibre diffraction methods.

This type of investigation could be improved in numerous ways. Larger crystal models would be better, if accompanied by significantly faster computers, or periodic boundary conditions could be incorporated into MM3. The MM3(92) program incorporates a test on the angle of hydrogen bonding, although the problem of premature energy termination still exists. The CRSTL program allows a much larger crystal size to be employed, for more rigorous calculations. However it does not allow the molecules themselves to optimize, crucial for the study of the relative stabilities of different forms.

References

1. vanderHart, D. L. and Atalla, R. H. *Macromolecules* 1984, **17**, 1465
2. Sugiyama, J., Vuong, R. and Chanzy, H. *Macromolecules* 1991, **24**, 4108
3. Honjo, G. and Watanabe, M. *Nature* 1958, **181**, 326
4. Chanzy, H., Imada, K. and Vuong, R. *Protoplasma* 1978, **94**, 299
5. Allinger, N. L., Yuh, Y. H. and Liu, J.-H. *J. Am. Chem. Soc.* 1989, **111**, 8551
6. Allinger, N. L., Raman, M. and Liu, J.-H. *J. Am. Chem. Soc.* 1990, **112**, 8293

- 7 Liu, J.-H. and Allinger, N. L. *J. Am. Chem. Soc.* 1989, **111**, 8576
- 8 French, A. D., Rowland, R. S. and Allinger, N. L. in 'Computer Modeling of Carbohydrate Molecules', (Eds French, A. D. and Brady, J. W.), ACS Symposium Series 430, ACS Books, Washington, DC, 1980, p 121
- 9 Wondcock, C. and Sarko, A. *Macromolecules* 1980, **13**, 1183
- 10 Stipanovic, A. J. and Sarko, A. *Macromolecules* 1976, **9**, 851
- 11 Sarko, A., Southwick, J. and Hayashi, J. *Macromolecules* 1976, **9**, 857
- 12 Gardiner, E. S. and Sarko, A. *Can. J. Chem.* 1985, **63**, 173
- 13 Aablo, A. and French, A. D. *Macromolecules* (in press)
- 14 Brown, G. M. and Levy, H. A. *Acta Crystallogr.* 1979, **B35**, 656
- 15 Chu, S. S. C. and Jeffrey, G. A. *Acta Crystallogr.* 1968, **B24**, 830
- 16 Kanters, J. A., Roelofs, G., Alblas, B. P. and Meinders, I. *Acta Crystallogr.* 1977, **B33**, 665
- 17 Sheldrick, B. *Acta Crystallogr.* 1977, **B33**, 3003
- 18 Brown, G. M. and Levy, H. A. *Acta Crystallogr.* 1973, **B29**, 790
- 19 Henricssal, B., Perez, S., Tvaroska, I. and Winter, W. T. in 'The Structures of Cellulose: Characterization of the Solid State', (Ed. Atallas, R. H.), ACS Symposium Series no. 340, ACS Books, Washington, DC, 1987, p 38
- 20 French, A. D., Mohous-Riou, N. and Perez, S. *Carbohydr. Res.* (in press)
- 21 CHEM-X is developed and distributed by Chemical Design Ltd, Oxford, England
- 22 Ferro, D. R. and Hermans, J. *Acta Crystallogr.* 1977, **A33**, 345
- 23 Aablo, A. and French, A. D. (unpublished)
- 24 Okano, T. and Koyanagi, A. *Biopolymers* 1986, **25**, 851
- 25 Dowd, M. K., French, A. D. and Reilly, P. J. *Carbohydr. Res.* 1992 (in press)
- 26 'CRC Handbook of Physics and Chemistry', 1983, p 12181
- 27 French, A. D., Schäfer, L. and Newton, S. Q. *Carbohydr. Res.* (in press)

10

Calculations of potential energy of the cellulose crystal structure

Alvo Asblöö, Aleksandr J. Pertsis* and R.-H. Mikelsaar** - Tartu University, Department of Experimental Physics, Tähe 4 Street, 202400 Tartu, Estonia. *Institute of Element-Organic Compounds, Vavilova 28 Street, GSPI, V-334, 117813 Moscow. **Tartu University, Institute of General and Molecular Pathology, Veski 34 Street, 202400 Tartu, Estonia.

ABSTRACT

Many aspects of cellulose crystal structure remain without satisfactory explanation. We have calculated objective functions based on diffraction intensities and potential energies of the cellulose II crystal using "rigid model" calculations. We used different geometries of the glucose rings and found that this does not shift the results of potential energy calculations very much. The second part of this work contains preliminary results of potential energy calculations for the cellulose Ia crystal structure. Models with the lowest energy are in the tg conformation. However, the best gg model has very complicated hydrogen bonding that includes two hydrogen bonds between sheets.

INTRODUCTION

In recent years there have been only a few attempts to determine the structure of cellulose by combined potential energy and X-ray diffraction calculations [1]. A typical X-ray diagram of cellulose contains only few dozen reflections, a number that is clearly insufficient to refine all atomic parameters by standard crystallographic methods. To increase the ratio of observations to refineable parameters we have used various stereochemical and packing constraints and non-bonded contacts calculated by the atom-atom potential method. For calculating the most

probable models of a crystal of cellulose we have minimized the objective function

$$F = U + W^*R^*$$

where U is the potential energy of the system, R^* is the crystallographic discrepancy factor based on the cellulose II intensity data from Kolpak et al [2]. The potential energy consists of

$$U = U_{\text{con}} + U_{\text{int}} + U_{\text{inter}}$$

in which U_{con} is the conformational energy of the monomer residues, U_{int} is the conformational energy between two successive residues along both of the two crystallographically distinct chains, U_{inter} is the intermolecular energy which includes non-bonded and H-bond energy between the atoms of different chains. Calculations have been executed in internal coordinates. There are two different extents of atomic adjustments that can be used to reduce F . The first possibility is to leave all bond lengths etc. flexible. The second way is to fix some structural components that do not vary much during minimization. This method allows us to decrease the number of variables. In earlier work [3] thirteen most probable models for the cellulose II crystal structure were calculated.

CALCULATIONS OF THE POTENTIAL ENERGY OF THE CRYSTAL STRUCTURE OF CELLULOSE II

To learn the importance of variable residue geometry, we have calculated all thirteen most probable models of cellulose II crystal structure [3] by the use of the "rigid model" method described in [3,4], applying five different initial glucose rings. We used β -glucose rings constructed by chiral inversion of residues found in cyclohexaamylose-KOAc/CHA2/, plantose/PLA/ and glucose-urea/GUR/ [5]. We also used reducing β -D-cellobiose/CBIO/ and β -D-glucose/ β GLU/ [6]. We presumed symmetry P2₁ for the chain of cellulose II. During minimization we fixed the bond lengths, bond angles and glucose ring geometry. We varied the torsional angles of the hydroxymethyl and three hydroxyl groups and the torsional and bond angles describing the junction between two successive monomer residues. We also varied rotations and shift of chains in the unit cell w cellulose II [2]. As for the most probable models, their variable parameters do not shift remarkably when the "different geometries are used. The only models that shifted were ones that ranked last in the initial preference list [3] (using "average Arnott-Scott" ring). When the model A11 was based on PLA and β GLU rings, it did not find the same energy minimum. Instead the A11 shifted to the A1 model. In Figs. 1 and 2 we present the value of the objective function and the potential energies of the thirteen most probable models of crystal structure II. It can be seen

seen that the models of lowest values of energy do not change remarkably. We deduce that "rigid" model calculations are quite correct for the initial calculations of cellulose structure. For more exact calculations it is reasonable to use "flexible" methods (as MM3) in which "rigid model" results can be used as initial variable sets.

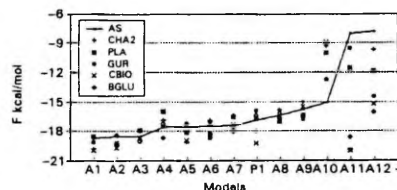


Figure 1. Objective functions of most probable models of cellulose II crystal. Glucose rings: AS - average Arnott-Scott; CHA2 - cyclohexamylose; PLA - plantose; GUR - gluco-urea; CBIO - cellobiose; BGLU - β -D-glucose. A1-A12, P1 - most probable models of the unit cell of cellulose II [3].

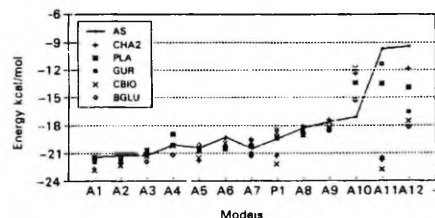


Figure 2. Potential energies of most probable models of the cellulose II crystal. For abbreviations see Fig. 1.

PRELIMINARY CALCULATIONS OF THE POTENTIAL ENERGY OF THE CELLULOSE I α CRYSTAL STRUCTURE

For a long time there were many questions about the structure of crystalline cellulose I. Several authors [7,8] tried to solve the problem, but it has up to now been unsolved. Later, using NMR data [9,10] it was found that the crystal of cellulose I consists of two phases: cellulose I α which must have triclinic symmetry, and cellulose I β which must have monoclinic symmetry. This year Sugiyama and collaborators [11] reported a new structure of the crystal of cellulose I. They used a new precision electron diffraction apparatus which allows them to apply a very narrow initial electron beam. They found that cellulose I α has a one-chain triclinic unit cell with parameters: $\alpha=117^\circ$, $\beta=113^\circ$, $\gamma=81^\circ$, $a=6.74\text{\AA}$, $b=5.93\text{\AA}$, $c=10.36\text{\AA}$. The aim of our work is to examine that unit cell from the energetic viewpoint. Unfortunately, X-ray diffraction data of phase I α do not exist yet. We have calculated the potential energy of the crystal of cellulose I α using "Arnott-Scott" glucose ring geometry for "rigid model" calculations. This time we presumed symmetry P1 for the chain of cellulose I α . We varied torsional angles of the two hydroxymethyl and six hydroxyl groups and the two torsional and bond angle between two successive units along the chain (Figure 3). Table 1 shows the 19 models having lowest energy for phase I α . The best "up" model (U1) has an energy of -19 kcal/mol in tg conformation. It is remarkable that at energies -14 to -16 kcal/mol there exist 6 "up" models with different hydrogen bonds. The best "down" model (D1) has an energy of -17 kcal/mol. The best gg model has an energy of -14 kcal/mol. It has very complicated hydrogen bonding. There are two hydrogen bonds between sheets along the a-axis.

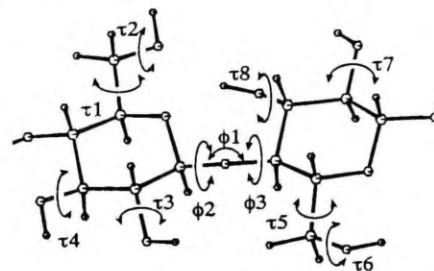


Figure 3. Variable torsional and bond angles in the chain of cellulose I α .

Table 1. The most probable models of the cellulose Ia triclinic unit cell.

U₊ - "up" models; D₋ - "down" models; τ_{ij} - variable bond and torsional angles (see Fig. 3); ϵ - angle describing chain rotation; τ - H-bonds in a chain; τ' - H-bonds in a sheet $\langle 110 \rangle$; τ'' - H-bonds between sheets $\langle 100 \rangle$.

| Energy | Bond | | | | | | | | | | | | | Def. | H-bonds |
|--------|-------|------|-----|------|------|-----|-----|-----|------|------|------|-----|-----|--|---------|
| 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | | |
| U1 | -19.5 | 71 | 166 | -52 | 175 | 118 | 93 | 146 | 73 | 166 | -52 | 175 | -77 | τ_{ij} O5...HO3 ¹¹ O2B...O4 ¹¹ | |
| U2 | -16.8 | 59 | 166 | -53 | 166 | 117 | 87 | 143 | 74 | 163 | -53 | 170 | -1 | O4B...O3 ¹¹ O5...HO3 ¹¹ O2B...O4 ¹¹ | |
| U3 | -16.3 | 69 | -61 | 166 | 166 | 119 | 93 | 146 | 72 | -63 | 170 | 171 | -72 | O4B...O3 ¹¹ O5...HO3 ¹¹ O11O2...HO4 ¹¹ | |
| U4 | -16.2 | 68 | 103 | -56 | 161 | 115 | 94 | 147 | 69 | 102 | -56 | 161 | -78 | τ_{ij} O5...HO3 ¹¹ O2B...O4 ¹¹ | |
| U5 | -15.7 | 64 | 68 | -79 | 177 | 116 | 95 | 146 | 65 | 57 | -79 | 177 | -79 | O5...HO3 ¹¹ O2B...O4 ¹¹ | |
| U6 | -15.0 | 69 | 34 | -54 | 73 | 114 | 94 | 146 | 71 | 15 | -54 | 73 | -78 | τ_{ij} O2B...O4 ¹¹ O3B...O6 ¹¹ | |
| U7 | -14.3 | 101 | 162 | -52 | 161 | 115 | 94 | 147 | 101 | 162 | -52 | 161 | -79 | O5...HO3 ¹¹ O4B...O3 ¹¹ O2B...O4 ¹¹ | |
| U8 | -14.3 | 50 | 108 | -55 | 166 | 119 | 93 | 143 | 70 | 109 | -56 | 171 | -6 | τ_{ij} O5...HO3 ¹¹ O2B...O4 ¹¹ O4B...O3 ¹¹ | |
| U9 | -14.2 | 53 | -53 | -150 | 175 | 119 | 83 | 138 | 72 | -50 | -150 | 177 | -1 | O5...HO3 ¹¹ O11O2...HO4 ¹¹ O4B...O3 ¹¹ | |
| U10 | -14.1 | -35 | -42 | -45 | 185 | 115 | 100 | 138 | -17 | -113 | 24 | 186 | -87 | τ_{ij} O5...HO3 ¹¹ O2B...O4 ¹¹ O4B...O3 ¹¹ O4B...O3 ¹¹ | |
| U11 | -13.6 | 76 | 172 | 183 | 167 | 117 | 94 | 148 | 76 | 172 | 183 | 167 | -78 | τ_{ij} O5...HO3 ¹¹ O4B...O3 ¹¹ O2B...O4 ¹¹ | |
| U12 | -13.1 | 73 | -49 | -156 | 176 | 117 | 95 | 141 | 74 | -49 | -157 | 176 | -78 | τ_{ij} O5...HO3 ¹¹ O11O2...HO4 ¹¹ O4B...O3 ¹¹ | |
| U13 | -12.9 | 46 | 61 | -67 | 175 | 118 | 84 | 140 | 74 | 67 | -69 | 177 | -3 | τ_{ij} O5...HO3 ¹¹ O2B...O4 ¹¹ O4B...O3 ¹¹ O11O2...HO4 ¹¹ | |
| U14 | -12.6 | -29 | 81 | 184 | 176 | 115 | 93 | 141 | -102 | 61 | 184 | 178 | 18 | τ_{ij} O5...HO3 ¹¹ O11O2...HO4 ¹¹ O2B...O4 ¹¹ | |
| U15 | -12.4 | 47 | 35 | -54 | 66 | 119 | 81 | 141 | 76 | 35 | -55 | 65 | -6 | τ_{ij} O2B...O4 ¹¹ O3B...O6 ¹¹ O4B...O3 ¹¹ | |
| U16 | -12.3 | 43 | -51 | -148 | 177 | 116 | 95 | 139 | 69 | -50 | -147 | 177 | -86 | τ_{ij} O5...HO3 ¹¹ O11O2...HO4 ¹¹ O4B...O3 ¹¹ O2B...O4 ¹¹ | |
| U17 | -12.3 | -156 | -61 | -37 | 169 | 117 | 94 | 147 | -129 | -61 | -36 | 176 | -78 | τ_{ij} O5...HO3 ¹¹ O2B...O4 ¹¹ O4B...O3 ¹¹ | |
| U18 | -12.2 | 187 | 175 | -112 | 181 | 117 | 97 | 128 | 192 | 171 | -114 | 181 | 25 | τ_{ij} O5...HO3 ¹¹ O4B...O3 ¹¹ O2B...O4 ¹¹ | |
| U19 | -12.1 | 100 | 177 | 174 | -175 | 115 | 96 | 142 | 94 | 174 | 171 | 173 | 3 | τ_{ij} O5...HO3 ¹¹ O4B...O3 ¹¹ O2B...O4 ¹¹ | |

Comparing these crystal energies with energies of a crystal of cellulose II which are calculated by the same method, it can be seen that cellulose Ia has slightly higher energy minima than cellulose II. It can be explained by the conversion of cellulose I into cellulose II. This conversion must be energetically profitable.

REFERENCES

1. Millane, R. P.; Narasiah, T. V.; *Polymer* **1989**, *30*, 1763.
2. Kolpak, F. J.; Weih, M.; Blackwell, J.; *Polymer* **1978**, *19*, 123.
3. Pertsin, A. J.; Nugmanov, O. K.; Marchenko, G. N.; Kitaigorodsky, A. I.; *Polymer* **1984**, *25*, 107.
4. Pertsin, A. J.; Kitaigorodsky, A. I.; *The Atom - Atom Potential Method Applications to Organic Molecular Solids*, Springer Ser. Chem. Phys., V43, Springer-Verlag, Berlin Heidelberg New-York London Paris Tokyo, 1987.
5. French, A. D.; Murphy, V. G.; *Carbohydrate Research* **1973**, *27*, 391.
6. Chu, S. S. C.; Jeffrey, G. A.; *Acta Cryst. B* **1970**, *26*, 1373.
7. French, A. D.; *Carbohydrate Research* **1978**, *61*, 67.
8. Gardner, K. H.; Blackwell, J.; *Biopolymers* **1974**, *13*, 1975.
9. Horii, F.; Hirai, A.; Kitamaru, R.; *Macromolecules* **1987**, *20*, 2117.
10. Vandertart, D. L.; Atalla, R. H.; *Macromolecules* **1984**, *17*, 1465.
11. Sugiyama, J.; Vuong, R.; Charzy, H.; *Macromolecules* **1991**, *24*, 4108.

Antiparallel molecular models of crystalline cellulose

R.-E. Mikalson and Alvo Aablos* - Tartu University, Institute of General and Molecular Pathology, 34 Veski Street, 202400 Tartu, Estonia. *Tartu University, Department of Experimental Physics, 4 Tähse Street, 202400 Tartu, Estonia.

KEYWORD: molecular modelling, structure of cellulose I.

ABSTRACT

Antiparallel molecular structure models are proposed for cellulose I having the same geometric parameters which were established by Sugiyama et al. [1] for cellulose α and β phases.

INTRODUCTION

Although the presence of two crystalline phases in native cellulose (I) was demonstrated by NMR spectroscopy in 1984 [2,3], the exact unit cell geometries of the α and β phases were established only recently [1]. Sugiyama et al. [1] interpret their experimental data according to a parallel molecular chain hypothesis.

DISCUSSION

The aim of the present paper is to show that antiparallel cellulose molecular structures may exist having the same unit cell dimensions. Molecular modelling was carried out by Tartu plastic space-filling atomic models [4]. The results of the modelling experiments allowed us to present following schemes of cellulose structure (Fig. 1-5).

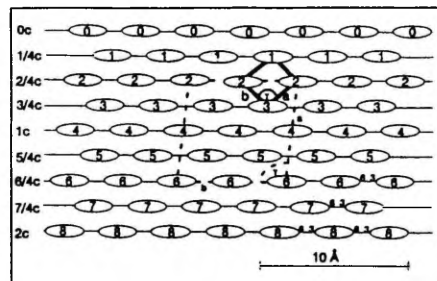


Figure 1. A structure of cellulose Ia phase according to Sugiyama et al [1]. Symbols: white ovals - parallel chains; a,b,c, γ - unit cell dimensions; numbers in ovals - the amount of c/4 shifts in the direction of c - axis of this unit cell; 6-3 - bonds O6H...O3.

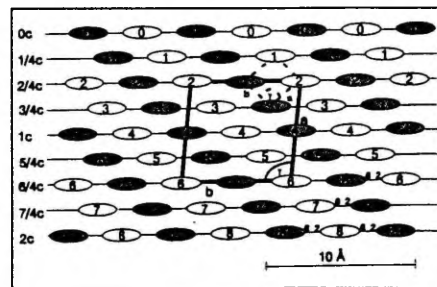


Figure 2. Antiparallel cellulose structure model "Ia-V". Symbols: white ovals - parallel chains; grey ovals - antiparallel chains; a,b,c, γ - unit cell dimensions; number in ovals - the amount of c/4 shifts in direction of c - axis of this unit cell; 6-2 - bonds O6-H...O2. Each chain(oval) contains two intra molecular bonds (O3-H...O5' and O2'-H...O6) and oxymethyl groups in tg or gg conformation.

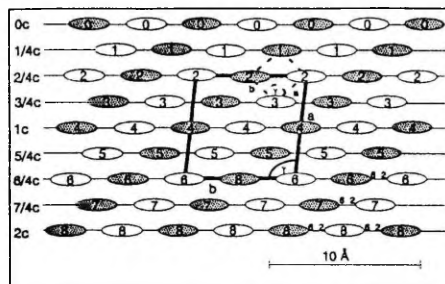


Figure 3. Antiparallel cellulose structure model "1a-1c". Symbols see Fig. 2.

The analysis of schemes presented in current paper allow us to suppose that antiparallel molecular structures of cellulose may have unit cell parameters corresponding to those given in the work of Sugiyama et al. [1]. Antiparallel structure models should be considered in the interpretation both of cellulose I and cellulose II investigations [5-7].

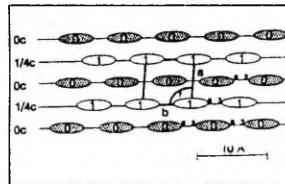


Figure 4. Antiparallel model of structure of cellulose 1B phase. Symbols see Fig. 1 and 2.

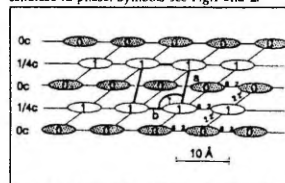


Figure 5. Antiparallel model of structure of cellulose II. Symbols see Fig. 1. and 2.

REFERENCES

1. Sugiyama, J., Vuong, R., Chanzy, H. *Macromolecules* 1991, 24, 4168.
2. Atalla, R. H., VanderHart, D. G. *Science* 1984, 223, 283.
3. VanderHart, D. G., Atalla, R. H. *Macromolecules* 1984, 17, 1465.
4. Mikelsaar, R. *Trends in Biotechnology* 1986, 6, 162.
5. Stipanovic, A. J., Sarko, A. *Macromolecules* 1976, 9, 851.
6. Kolpak, J. F., Weih, M., Blackwell, J. *Polymer* 1978, 19, 123.
7. Takahashi, Y., Matsunaga, H. *Macromolecules* 1991, 24, 3968.

in press.

Preliminary Potential Energy Calculations of Cellulose I α Crystal Structure.

Alvo Aabloo

Tartu University, Department of Experimental Physics, T  he 4 Street, 202400 Tartu, Estonia

*Alfred D. French**

Southern Regional Research Center, U.S. Department of Agriculture, 1100 Robert E. Lee Blvd., P.O. Box 19687, New Orleans, LA 70179, U. S. A.

(Received: January 1993)

SUMMARY:

Packing energy calculations were used to evaluate various models of I α cellulose, based on unit cell dimensions proposed by Sugiyama *et al.* Both a rigid-ring method, PLMR, and a full-optimization molecular mechanics technique, MM3, were used. The model found to be best with both methods was packed "up" (the z coordinate of O(5) is greater than that of C(5)); O(6) atoms were in *tg* positions, forming sheets of hydrogen bonded chains. With the PLMR program, the energy of the best model was almost 3 kcal/mol lower than the second best model. The MM3 studies also showed substantially higher energies for the alternative models. Also, some alternative PLMR models had substantially higher atomic movement during MM3 optimization.

Several years ago, VanderHart and Atalla found that all crystalline native cellulose I is composed of two phases ¹⁾. More recently, Sugiyama *et al.* characterized those two phases in *Microdictyon tenuius* with selected area electron diffraction techniques ²⁾. Two unit cells were resolved, a monoclinic,

two-chain cell and a triclinic, one-chain cell. When roughly equivalent amounts of both phases diffract simultaneously, the resulting diffraction pattern is identical to the pattern first indexed as an eight-chain cell by Honjo and Watanabe ³⁾. Thus, the eight-chain cell is apparently an artifact of the previous diffraction techniques. Previous structural studies were based on the assumption that cellulose I is only a single phase. Since those patterns contained information from two phases, new studies of cellulose I structure are needed.

However, x-ray diffraction data are not yet available, so it is useful to propose a crystal structure based on modeling methods. The triclinic, one-chain unit cell is especially conducive to structural studies using packing energy because the variety of possible structures is minimal. All proposals must be based on parallel chains, and each chain must be identical with its neighbors. This communication reports the results of two different types of packing energy calculations using various models of I α cellulose. All studies used the cell parameters $a=6.74 \text{ \AA}$, $b=5.93 \text{ \AA}$, $c=10.36 \text{ \AA}$, $\alpha=117^\circ$, $\beta=113^\circ$ and $\gamma=81^\circ$, as published by Sugiyama *et al.* ²⁾ The present work shows that reasonable models exist for those published unit cell dimensions. Our work is also part of a larger project to screen, with molecular mechanics, structures first proposed based on fiber diffraction studies.

The first method used to select likely structures was based on the PLMR program developed by Pertsin and Kitaigorodsky ^{4,5)}. It kept the Arnott-Scott ⁶⁾ residues used in the starting structures rigid except for orientation of the primary and secondary alcohol groups (see Fig. 1). The chain was free to rotate about its axis in the unit cell, and the torsion angles and bond angles at the glycosidic linkages were allowed to vary, consistent with the P1 space group. In the second method, model cellulose crystals were built of seven cellotetraose molecules, placed in a manner similar to hexagonal close packing, but with the chains arranged according to the I α unit cell dimensions (see Fig. 2). The starting structures were based on the best models from the PLMR work. The model crystals were optimized with the molecular mechanics program, MM3 ^{7,8)}, allowing all atoms to seek positions of minimum energy. In the MM3 optimizations, there were no restrictions of any kind on atomic movement other than those of the relatively complex MM3 force field. Unlike the PLMR program, MM3 requires the neighboring molecules to be explicitly included. Because MM3, as distributed, has a limit of 700 atoms, we used a relatively small model, similar to the initial nucleus that eventually grows into a crystal. For this reason, cellotetraose molecules were chosen to represent the very long cellulose chain molecules.

In the PLMR calculations, the potential energy consisted of intramonomeric, junction and intermolecular values. The intramonomeric term consisted of torsional potentials for the side group rotations. The junctional energy consisted of torsional potentials for the rotations, bending potentials for the angles, and intermonomeric hydrogen bonding and van der Waals terms. The intermolecular term consisted of hydrogen bonding and van der Waals terms. The starting models used in the PLMR calculations were based on an assumption that the cellulose chains would be centered in the unit cell and aligned approximately with the 110 plane as shown in Fig. 2; chain rotation was also one of the refined parameters. Surveys indicated the likely orientations of the side groups, and about 500 of the most likely structures were energy-minimized, based on Powell's quasi-Newton algorithm.

Thirty five of the 500 models optimized to give PLMR energies lower than -10 kcal/mol, of which the 10 lowest are in Table 1. The best "up" model (U1) has an energy of -19 kcal/mol. This value can be compared with the value of -21 kcal/mol for a model of cellulose II calculated with the same method. (These arbitrary energies are for a cellobiose unit in either case).

Although U1 is lower in energy than the others in Table 1 by about 3 kcal/mol, there are six models within -14 to -16 kcal. The best "down" model (D1) has an energy of -17 kcal/mol. The best nine models have O(6) in

the *tg* conformation. The best *gg* model has an energy of -14 kcal/mol, and the best *gt* model is even higher in energy at -12 kcal/mol. The *gg* models all have complicated hydrogen bonding systems (see Table 2).

In the MM3 studies, the energy optimization option was used, as was a dielectric constant of 4. This value was also chosen for studies of crystalline amides, polypeptides and proteins.¹⁰ Hydroxyl hydrogen atoms on O(1) and O(4) on the ends of the tetramers were manually adjusted to positions that gave low energy and permitted minimization to reach a normal termination. The mean atomic movement (Table 3) after a least squares fit was computed using CHEM-X⁹⁾. These mean atomic movements were calculated both with the hydrogen atoms included and excluded. They show the differences between the structures as determined by PLMR and the final MM3 structures. The energy of the model crystal is affected by all of the intra- and as well as the inter-molecular interactions, so the MM3 calculations are potentially more valid than rigid residue calculations that do not contain intra-residue bending and stretching energies. The mean atomic movement resulting from MM3 optimization shows the extent of agreement of the two methods.

Table 3 shows that U1 is superior according to the energy values from both the MM3 and the PLMR calculations. The mean atomic movements for U1 are also lower than for some of the other structures. Coordinates for the

MM3-optimized U1 model are in Table 4, and the optimized values of the PLMR variables are given in Table 1. Because of slight distortions within the MM3 model, the value of the chain rotation, ϵ , loses its meaning and was not calculated, despite the freedom for the chain to rotate during MM3 optimization. The best *gg* model, U7, is clearly worse than the others, based on the movement criterion. U5, with unusually low movement but high energy, appears to have been trapped in a high-energy local minimum. Two of the structures determined with PLMR, U3 and U4, were not stable in the MM3 force field, despite low movements of the non-hydrogen atoms, because hydrogen bonds were formed, giving structures similar to U1. The relatively high movements of hydrogen atoms show that the structure changed to form the hydrogen bonds (Table 3). Despite the similarity with U1 for the U3 and U4 chains that could form hydrogen bonds, the hydroxyl groups on the other chains in the models remained in their initial conformations, hence the high energies of the final models.

The MM3 energy values in Table 3 can be compared with our values of 200 kcal/mol¹¹⁾ calculated for Woodcock and Sarko's ramie cellulose structure¹²⁾ and 176 kcal/mol for Stipanovic and Sarko's structure of cellulose II¹³⁾. The movement values can be compared with non-hydrogen movements that we calculated for a similar model crystals of small carbohydrates (0.09 to

References.

1. D. L. VanderHart, R. H. Atalla, *Macromolecules* **17**, 1465 (1984)
2. J. Sugiyama, R. Vuong, H. Chanzy, *Macromolecules* **24**, 4108 (1991)
3. G. Honjo, M. Watanabe, *Nature* **181**, 326 (1958)
4. A. J. Pertsin, A. I. Kitaigorodsky, "*The Atom - Atom Potential Method Applications to Organic Molecular Solids*", Springer Ser. Chem. Phys., V43, Springer-Verlag, Berlin, 1987.
5. A. J. Pertsin, O. K. Nugmanov, G. N. Marchenko, A. I. Kitaigorodsky, *Polymer* **25**, 107 (1984)
6. S. Arnott, W. E. Scott, *J. Chem. Soc. Perkin II*, 324 (1972)
7. N. L. Allinger, Y. H. Yuh, J.-H. Lii, *J. Amer. Chem. Soc.* **111**, 8551 (1989)
8. N. L. Allinger, M. Rahman, J.-H. Lii, *J. Amer. Chem. Soc.* **112**, 8293 (1990)
9. CHEM-X is developed and distributed by Chemical Design Ltd., Oxford, England.
10. J.-H. Lii and N. L. Allinger, *J. Comput. Chem.* **12**, 186 (1991)
11. A. D. French, D. P. Miller and A. Aabloo, *Int. J. Biol. Macromol.* **15**, 30 (1993)
12. C. Woodcock, A. Sarko, *Macromolecules* **13**, 1183 (1980)
13. A. J. Stipanovic, A. Sarko, *Macromolecules* **9**, 851 (1976)

0.209 Å)¹¹. Based on continuing work with these small model crystals, we now believe that movements of much less than 0.20 Å correspond to structures that are not fully optimized because of difficulties¹¹ with block-diagonal MM3 optimizations. The value of energy for U1 is higher than the cellulose II energy, as would be expected. If U1 and the cellulose II structure are correct, then the proposed ramie structure may not be correct for cellulose I β , since annealing cellulose I α converts it to I β . Therefore, the energy of I β should be lower than the I α energy. Still, the closeness of the energy values for several models from both the PLMR and MM3 methods indicate that the conclusive determination of the crystal structures of these new phases will be difficult.

Table 1. Most probable PLMR models of the cellulose Ia triclinic unit cell.

U - "up" models; D - "down" models: Variable angles: $\tau_1 = \text{C4-C5-C6-O6}$, $\tau_2 = \text{C5-C6-O6-H}$,
 $\tau_3 = \text{C1-C2-O2-H}$, $\tau_4 = \text{C2-C3-O3-H}$, $\tau_5 = \text{C4'-C5'-C6'-O6'}$, $\tau_6 = \text{C5'-C6'-O6'-H'}$,
 $\tau_7 = \text{C1'-C2'-O2'-H}$, $\tau_8 = \text{C2'-C3'-O3'-H'}$, $\phi_1 = \text{C1-O4'-C4'}$, $\phi_2 = \text{O5-C1-O1-C4'}$,
 $\phi_3 = \text{C1-O1-C4'-C5'}$ (see Fig. 1); ϵ - angle describing chain rotation.

| Energy | | | | | | | | | | | | | | |
|-----------------------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|------------|----------|
| Model | kcal/mol | τ_1 | τ_2 | τ_3 | τ_4 | ϕ_1 | ϕ_2 | ϕ_3 | τ_5 | τ_6 | τ_7 | τ_8 | ϵ | O6 Conf. |
| U1 | -19.5 | -71 | -116 | 52 | -175 | 118 | -93 | -146 | -73 | -166 | 52 | -175 | 77 | tg |
| D1 | -16.8 | -59 | -166 | 53 | -168 | 117 | -87 | -143 | -74 | -163 | 53 | -170 | 1 | tg |
| U2 | -16.3 | -69 | 61 | 164 | -168 | 119 | -92 | -148 | -72 | 63 | 170 | -171 | 72 | tg |
| U3 | -16.2 | -68 | -102 | 56 | -161 | 115 | -94 | -147 | -69 | -102 | 56 | -161 | 78 | tg |
| U4 | -15.7 | -64 | -60 | 79 | -177 | 116 | -95 | -146 | -66 | -57 | 79 | -177 | 79 | tg |
| U5 | -15.0 | -69 | -34 | 54 | -72 | 116 | -94 | -146 | -71 | -35 | 54 | -72 | 78 | tg |
| U6 | -14.3 | -101 | -162 | 52 | -161 | 115 | -94 | -147 | -101 | -162 | 52 | -161 | 79 | tg |
| D2 | -14.3 | -50 | -108 | 55 | -168 | 119 | -83 | -143 | -70 | -109 | 56 | -171 | 6 | tg |
| D3 | -14.2 | -53 | 53 | 150 | -175 | 119 | -83 | -138 | -72 | 50 | 150 | -177 | 1 | tg |
| U7 | -14.1 | 35 | 62 | 45 | -185 | 115 | -100 | -138 | 17 | 113 | -24 | -186 | 97 | gg |
| U1 (after MM3 optimization) | | | | | | | | | | | | | | |
| | | -65 | -167 | 61 | 155 | 116 | -92 | -150 | -75 | -158 | 60 | -177 | -- | tg |

Table 2. Hydrogen bonds of ten most probable PLMR models of the cellulose Ia crystal.

U_ - "up" models; D_ - "down" models; ¹ - H-bonds in a chain;
² - H-bonds in sheets parallel to the <110> planes; ³ - H-bonds
between sheets (parallel to the <100> planes); ⁴ - H-bonds which
were found by MM3.

| | | | | |
|----|----|-----------------------|--------------------------|---|
| U1 | tg | O5..HO3' ¹ | O2H..O6' ¹ | O6H..O3 ² |
| D1 | tg | O5..HO3' ¹ | O2H..O6' ¹ | O6H..O3 ² |
| U2 | tg | O5..HO3' ¹ | O1;O2..HO6' ¹ | |
| U3 | tg | O5..HO3' ¹ | O2H..O6' ¹ | O6H..O3 ⁴ |
| U4 | tg | O5..HO3' ¹ | O2H..O6' ¹ | O6H..O3 ⁴ |
| U5 | tg | O2H..O6' ¹ | O3H..O6 ² | |
| U6 | tg | O5..HO3' ¹ | O6H..O2;O3 ² | |
| D2 | tg | O5..HO3' ¹ | O2H..O6' ¹ | O6H..O3 ² |
| D3 | tg | O5..HO3' ¹ | O1;O2..HO6' ¹ | |
| U7 | gg | O5..HO3' ¹ | O2H..O6' ¹ | O6H..O5 ³ O6H..O2 ³ |

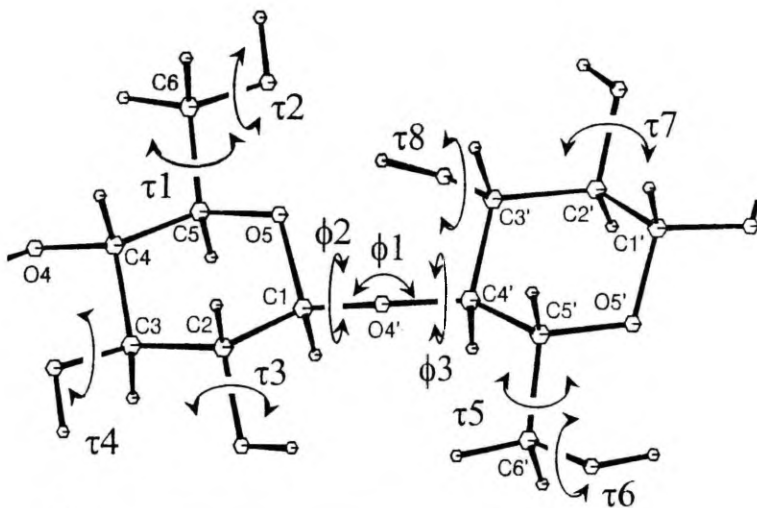
Table 3. The results of PLMR energy (kcal/mol of cellobiose residues) and MM3 energy (kcal/mol of seven tetraose units) calculations. For MM3, PLMR models were used as input. Mean deviations (Å) are calculated with hydrogens and without hydrogens.

| Model | Energy | | Mean atom movement | | Conf. | Remarks. |
|-------|--------|-----|--------------------|----------------|---------|-------------------|
| | PLMR | MM3 | with H's | without H's | | |
| U1 | -19.5 | 185 | .175 | .106 | up tg | |
| D1 | -16.8 | 206 | .158 | .108 | down tg | |
| U2 | -16.3 | 222 | .178 | .148 | up tg | |
| U3 | -16.2 | 218 | .177 | .092 | up tg | O6H..O3 formation |
| U4 | -15.7 | 215 | .193 | .098 | up tg | O6H..O3 formation |
| U5 | -15.0 | 234 | .056 | .035 | up tg | |
| U6 | -14.3 | 210 | .257 | .163 | up tg | |
| D2 | -14.3 | 206 | .232 | .158 | down tg | |
| D3 | -14.2 | 213 | .277 | .234 | down tg | |
| U7 | -14.1 | 229 | .451 | .384 | up gg | |

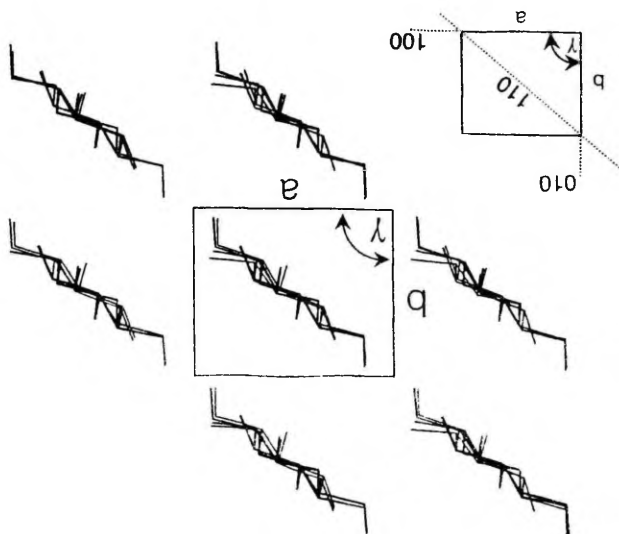
Table 4. Cartesian coordinates of the best model (U1) of cellulose I α .
The atoms are from the central two residues in the middle of the minicrystal.

| Name | x (Å) | y (Å) | z (Å) |
|----------|----------|----------|---------|
| 1 O4 | .00000 | .00000 | .00000 |
| 2 C4 | .21142 | .75417 | 1.20317 |
| 3 C5 | .72797 | -.18670 | 2.29622 |
| 4 O5 | .81127 | .54182 | 3.51681 |
| 5 C1 | -.46448 | .97228 | 3.98600 |
| 6 C2 | -1.02514 | 1.97530 | 2.98240 |
| 7 C3 | -1.14734 | 1.32907 | 1.60488 |
| 8 C6 | 2.13877 | -.69315 | 2.01594 |
| 9 O4' | -.30160 | 1.67408 | 5.19677 |
| 10 O2 | -2.34490 | 2.31506 | 3.42432 |
| 11 O3 | -1.50426 | 2.35082 | .66168 |
| 12 H-C4 | .94094 | 1.57657 | 1.02398 |
| 13 H-C5 | .04336 | -1.05521 | 2.41776 |
| 14 H-C1 | -1.15372 | .11065 | 4.10344 |
| 15 H-C2 | -.38335 | 2.88275 | 2.92769 |
| 16 H-C3 | -1.92144 | .52822 | 1.62421 |
| 17 O6 | 2.14321 | -1.50498 | .83483 |
| 18 H-C6a | 2.83975 | .15639 | 1.88709 |
| 19 H-C6b | 2.49779 | -1.28948 | 2.88036 |
| 20 H-O6 | 2.97128 | -1.96758 | .80196 |
| 21 H-O2 | -2.26797 | 2.71555 | 4.28199 |
| 22 H-O3 | -1.51712 | 1.95128 | -.19857 |
| 23 C4' | -.39931 | .89310 | 6.38669 |
| 24 C5' | -.94217 | 1.83149 | 7.47428 |
| 25 O5' | -.98538 | 1.11979 | 8.70779 |
| 26 C1' | .31199 | .75252 | 9.17576 |
| 27 C2' | .91283 | -.22900 | 8.17675 |
| 28 C3' | .99336 | .39823 | 6.78732 |
| 29 C6' | -2.37821 | 2.28419 | 7.20241 |

| | | | |
|-----------|----------|----------|---------|
| 30 O2' | 2.25134 | -.50458 | 8.60715 |
| 31 O3' | 1.41171 | -.63944 | 5.88912 |
| 32 H-C4' | -1.10106 | .03922 | 6.26071 |
| 33 H-C5' | -.28930 | 2.72653 | 7.58103 |
| 34 H-C1' | .96555 | 1.64231 | 9.28895 |
| 35 H-C2' | .31372 | -1.16602 | 8.13966 |
| 36 H-C3' | 1.72346 | 1.23868 | 6.78605 |
| 37 O6' | -2.40745 | 3.25978 | 6.15041 |
| 38 H-C6a' | -3.01741 | 1.41961 | 6.92744 |
| 39 H-C6b' | -2.80889 | 2.72817 | 8.12550 |
| 40 H-O6' | -3.20988 | 3.76025 | 6.23720 |
| 41 H-O2' | 2.19927 | -.88692 | 9.47471 |
| 42 H-O3' | 1.50327 | -.25602 | 5.02697 |



1. A unit of the cellulose chain showing the numbering of the ring and glycosidic atoms and the torsion and bond angle variables of the PLMR program.



2. Model 1/1, after MM3 minimization, showing the unit cell parameters a , b and γ . The view is down the chain axes of the seven tetramers. Intermolecular hydrogen bonding occurs in the 110 planes of the crystal (the planes are defined in inset).

Studies of Crystalline Native Celluloses Using Potential Energy Calculations.

Running Title: Model Cellulose Ia and Ib

Alvo Aabloo^{*} University of Tartu, Institute of Experimental Physics and Technology, Tähe 4 Street, EE2400 Tartu Estonia; **Alfred D. French** Southern Regional Research Center, U.S. Department of Agriculture, 1100 Robert E. Lee Blvd., P. O. Box 19687, New Orleans, LA 70179, U. S. A; **Raik-Hiio Mikelsaar** University of Tartu, Institute of General and Molecular Pathology, Veski 35 Street, EE2400 Tartu, Estonia; **Aleksandr J. Pertsin** Institute of Element-Organic Compounds, Vavilova 28 Street, GSP1, V-334, 117813 Moscow, Russia.

Keywords: Cellulose I, molecular mechanics, crystal structure, phase change, modelling

Abstract

Energies for various trial packing arrangements of unit cells for the Ia and Ib phases of native cellulose discovered by Sugiyama et al. were evaluated. Both a rigid-ring method, PLMR, and the full-optimization, molecular mechanics program, MM3(90) were used. For both phases, the models that had lowest PLMR energy also had the lowest MM3 energy. Both have the chains packed "up", O-6's in *tg* positions, and the same sheets of hydrogen bonded chains. The Ib structure is essentially identical to the structure proposed previously for ramie cellulose by Woodcock and Sarko. It is also the same as the best parallel model previously proposed that was based on the X-ray data of Mann, Gonzalez and Wellard, once the various unit cell conventions are considered. Also, the energies from both methods for all three celluloses, Ia, Ib and II are in the order that rationalizes their relative stabilities.

Introduction

The two main forms of cellulose are I, the major native type, and II, which occurs after mercerization or regeneration. A proposal for the solid-state conversion from I into cellulose II, including different intermediates, is described elsewhere (Nishimura et al., 1991, Nishimura and Sarko, 1991). Some years ago, it was discovered (VanderHart and Atalla, 1984, Atalla and VanderHart, 1984) that native cellulose occurs mostly as combinations of the two phases, Ia and Ib. Depending on the sample, these phases are in different ratios. For example, the ratio Ia/Ib was estimated to be 65%/35% for *Valonia* cellulose (VanderHart and Atalla, 1984). Initially, electron diffraction patterns from this mixture were indexed based on an 8-chain unit cell (Honjo and Watanabe, 1958). More recently, however, electron diffraction from small selected areas of the same microfibril showed two patterns that were interpreted (Sugiyama et al., 1991) to have, respectively, one-chain, triclinic and two-chain monoclinic unit cells. One of the most important conclusions from the finding of a 1-chain unit cell is that the chains must have a parallel orientation, i.e. the reducing ends of the cellulose molecules must all be at the same end in a given microfibril. Because both phases are found within the same microfibril in that work, the conclusion of parallel packing for Ib is almost inescapable.

Several decades ago, Rånby (1952) showed that cellulose II is slightly more stable than cellulose I, by about 2 cal/gram or 0.3 kcal/mol of glucose residues. Because hydrothermal annealing converts mixtures of Ia and Ib into pure Ib (Horii et al., 1987), it is thought that Ib is more stable than Ia. As X-ray diffraction intensity data for the pure phases of native cellulose are yet not available, it

is reasonable to propose structures based on energy calculations. This was especially so for $I\alpha$, because of the limited number of variables that must be considered for a 1-chain model, and our preliminary results for $I\alpha$ were already published (Aabloo and French, in press). In the present work we have studied various possible model structures for the $I\alpha$ and $I\beta$ celluloses, based on the published unit cell dimensions and on calculated packing energy. These results are then compared with other work.

Methods.

All of our computer models consisted of parallel chains and were based on published unit cell dimensions (Sugiyama et al., 1991). For $I\alpha$, $a = 6.74 \text{ \AA}$, $b = 5.93 \text{ \AA}$, $c = 10.36 \text{ \AA}$, $\alpha = 117^\circ$, $\beta = 113^\circ$ and $\gamma = 81^\circ$, and for $I\beta$, $a = 8.01 \text{ \AA}$, $b = 8.17 \text{ \AA}$, $c = 10.36 \text{ \AA}$, $\alpha = 90^\circ$, $\beta = 90^\circ$ and $\gamma = 97.3^\circ$. Space group P1 was used for the $I\alpha$ form, and the cellulose $I\beta$ chain models conformed to space group P2₁. We used two different methods to calculate packing energies, PLMR and MM3. The PLMR (derived from a word *polymer*) program, described by Pertsin et al. (1984) and Pertsin and Kitaigorodsky (1987), uses a "rigid-ring" strategy. In PLMR models of the cellulose chains, most of the intraresidue parameters of the glucose rings are kept at average values reported by Arnott and Scott (1972). Only the exocyclic torsion angles τ_i (see Fig. 1) that locate the hydroxymethyl and three hydroxyl groups were varied. The chains were free to rotate about their axes in the unit cell. The angles φ_i describe those rotations. They are defined as the angles formed by vectors λ (see Fig. 2) and projection of the unit cell axes $a' = a \cdot \cos(\gamma - \pi/2)$. For $I\alpha$, the torsion angles and bond angles at the glycosidic linkages were allowed to vary. For $I\beta$, the monomer residues were linked to form the chain with the variable virtual bond method (Zugenmaier and Sarko, 1980). Keeping 2₁ symmetry requires an additional parameter that characterizes the chain conformation (see Fig. 2). The angle δ describes rotation of the monomer residue about the O4--O4' virtual bond. It is defined by three vectors λ , μ , and ν . The vector λ is perpendicular to the c-axis, μ is along the O4-C4 bond, and ν is along the O4--O4' virtual bond. The angle δ is the dihedral angle between the plane $\mu\nu$ and the plane $\nu\lambda$. In the variable virtual bond method, the conformational parameters at the junction of two successive residues are not independent variables, but are implicit functions of h (the rise per monomer along the c-axis), δ and the residue conformation. For the two-chain unit cell, one more parameter, s , characterizes the vertical shift, or stagger, of the central chain. In all, there were 12 independent variables to be minimized for $I\alpha$ and 13 for $I\beta$. One binary variable parameter defines the chain direction as "up" or "down". A chain is "up" if the z coordinate of O-5 is greater than the z-coordinate of C-5.

The PLMR potential energy U includes the intramonomer energy, junction energy, and intermolecular energy:

$$U = U_{\text{mon}} + U_{\text{junct}} + U_{\text{inter}}$$

in which U_{mon} includes the torsional and nonbonded contributions from dihedral angles and nonbonded contacts that are influenced by variation of the monomer parameters τ_i . Briefly, U_{junct} includes contributions from the bending of the glycosidic bond angle, from the torsional angles at the junction and from non-bonded contacts between two successive residues. U_{inter} includes nonbonded interactions between the atoms belonging to different chains, consisting of hydrogen bonding and van der Waals terms. Previously published formulas and the values of parameters are used to calculate bond angle energies (Pertsin and Kitaigorodsky, 1987). A Morse-like equation is used for hydrogen bond modelling and a 6th order exponential potential function is used for van der Waals calculations (Pertsin and

Kitaigordsky, 1987). As PLMR uses continuous chains, no electrostatic interactions were considered. Calculations were executed in internal coordinates. PLMR uses periodic boundary conditions to calculate the interchain energies.

Model cellulose crystals for the MM3(90) (Allinger et al., 1989, Allinger et al., 1990) molecular mechanics calculations were built of seven cellobiose molecules which represented the very long cellulose chain molecules. They were placed in a manner similar to hexagonal close packing, but conformed to the unit cell dimensions. Terminating O-4 and O-1''' hydrogens were moved manually to positions that gave low energy. The dielectric constant was 4.0, a value used for crystalline amides, polypeptides and proteins (Lii and Allinger, 1991). Starting structures in the MM3 calculations were taken from the best models from the PLMR process. The model crystals were optimized with MM3(90) with no restrictions such as unit cell dimensions or space group. All atoms were allowed to seek positions of minimum energy. The MM3 minimization process should be more valid than rigid-residue calculations because the molecule can relax and adapt to the local environment. Also, the energies calculated by MM3 are likely to be fairly accurate; the standard deviation in calculated heat of formation for 40 isolated alcohols and ethers was 0.38 kcal/mol, although energies for the minicrystals may be less accurate. On the other hand, MM3 takes considerable computer time to optimize a miniature crystal. PLMR rapidly scanned hundreds of trial structures and maintained the unit cell dimensions precisely.

Using these two minimization processes we got most probable models for both crystalline phases of cellulose. To learn the differences between the structures determined by PLMR and by MM3, we computed mean atomic movement using a routine in CHEM-X, developed and distributed by Chemical Design, Ltd, Oxford, U. K. These atomic movements were calculated with the hydrogen atoms included and excluded. The mean atomic movement indicates the degree of stability of the PLMR models in the MM3 force field. Some lattice expansion in the miniature crystal models is expected, because they are small, and do not fully account for long-range forces present in the somewhat larger actual microfibrils. However, the movements for the best cellulose models are somewhat smaller than computed for fully optimized similar models of small carbohydrate molecules (about 0.2 Å). This is because the averaged molecular geometries in the PLMR starting models are not changed as much by MM3 optimization as are structures determined individually by diffraction methods. Also, the crystals of the small molecules are much larger than cellulose microfibrils, and their long range forces result in slightly greater lattice compression.

The starting models for PLMR were developed with Tartu space-filling models (Mikelsaar, 1986) and computer graphics modeling. To avoid missing some reasonable energy minima we built several models in the regions of conformational space where we found structures having low energies. We also tried many models in other regions. The initial sets of starting models consisted of about five hundred models. "Up" and "down" starting models were constructed independently for both polymorphs α and β .

Results and discussion

Cellulose Ia

The models having the six lowest PLMR energies for Ia are presented in Table 1. The best model, U_a1, has an energy of -19.5 kcal/mol of cellobiose residues and is packed "up". The energy value can be compared with the value of -21.4 kcal/mol for a model of cellulose II calculated with same method (Pertsin et al., 1984). The best "down" model has an energy of -16.8 kcal/mol and is in second place. All six models have O-6 in the *tg* conformation. The best *gg* model had an energy of -14.1 kcal/mol and the best *gt* model had an energy of -12.3 kcal/mol. The hydrogen bonds of these six models are listed in Table 3. The *gg* and *gt* models had very complicated hydrogen bonding systems.

The structures having the lowest PLMR energies were then minimized using MM3. The best model, U_a1, kept its leading position, but some other models changed places (Table 3). Table 3 also presents mean atomic movements. The structure U_a1 preserves its conformation after MM3 minimization. Its MM3 energy is 185 kcal/mol for the assembly of 28 glucose residues. The structures U_a3 and U_a4 were not stable in the MM3 force field, but were trapped during PLMR minimization because of the use of rigid residues. These models optimized to structures similar to U_a1 in the MM3 calculations. The coordinates of atoms in the PLMR-minimized Ia structure are in Table 4.

Cellulose Ib

The models with the seven lowest PLMR energies for Ib are shown in Table 2. The best model, U_b1, has an energy of -19.9 kcal/mol. It is packed "up", and has O-6 in *tg* positions. The best *gg* and *gt* models had energies higher than -12 kcal/mol. Model U_b2 has the same hydrogen bonding scheme (Table 3) and chain conformation as U_b1, and differs only in the shift of the central chain.

The results of minimizing these seven structures with MM3 are shown in Table 3. As with the Ia studies, the model best by PLMR calculations also gave the lowest MM3 energy for Ib, 182 kcal/mol. This was slightly lower than for the Ia model and slightly higher than for cellulose II. Also shown are the mean atomic movements. It seems that models U_b2, U_b3 and U_b4 were trapped in local minima in the PLMR calculations. Structures U_b2 and U_b4 had high movements of oxygen atoms during MM3 minimizations. Thus, those PLMR structures were substantially disrupted by MM3 optimization. As in the PLMR work on Ia, the U3 structure did not have any intersheet hydrogen bonds. MM3 minimization did result in such hydrogen bonds for the U3 model, showing that these structures were caught in local minima in the PLMR studies and are not really stable. Structure U_b6 also formed new, intersheet hydrogen bonds during MM3 minimization.

Structure U_b1 is essentially identical to the structure proposed by Woodcock and Sarko for ramie cellulose I on the basis of X-ray fiber diffraction data and the limited force field incorporated in the PS-79 software (Woodcock and Sarko, 1980). Our previously published (French et al., 1993) value of 201 kcal/mol for the MM3 energy of their structure is in error because our model was inadvertently an incorrect representative of their structure. In our incorrect model, the central chain was translated in the opposite direction, causing the high energy. The U_b1 structure is also the same as the best

parallel model proposed (Pertsin et al., 1986) based on the PLMR program and the fiber diffraction data of Mann et al., (1958), once adjustments are made for the differences in unit cell conventions (French and Howley, 1989). The coordinates of atoms in the PLMR model of $I\beta$ are in Table 5.

Despite the difficulties with the application of MM3 to miniature crystals of cellulose that were documented in earlier (French et al., 1993), both minimization procedures agreed on the most probable structures for $I\alpha$ and $I\beta$ cellulose, models $U_{\alpha}1$ and $U_{\beta}1$. PLMR energies for cellulose $I\alpha$, $I\beta$ and II were -19.5, -19.9 and -21.4 (Pertsin et al., 1984) kcal/mol of cellobiose units. MM3(90) energies for the same series were 185, 182 and 181 kcal for the 7-chain minicrystals. (To obtain 181 kcal/mol for II, both model structures of cellulose II must be averaged, i.e., the model composed of four up gt chains and three down tg chains must be balanced by the model with four up tg chains and three down gt chains.) These trends of relative energies agree well with experiment. The energies of both phases of cellulose I are slightly higher than the energy of cellulose II, as expected from Rånby's experimental difference of 0.6 kcal/mol of cellobiose units.

The energies of $I\alpha$ and $I\beta$ are very close. This may explain why these celluloses coexist in a single microfibril. Both models have the same hydrogen bonding schemes. The conversion from $I\alpha$ to $I\beta$ may take place by a vertical shifting of sheets. During a modeled shift of sheets, without change in the lateral dimensions of the unit cell, PLMR calculations gave a barrier of 9 kcal/mol. Fully optimized structures with annealing conditions would probably give a barrier with lower energy.

The finding of the same hydrogen bonds in both systems conflicts with the conclusions of Wiley and Atalla (1987), who found that the hydrogen bonds must change during conversion from $I\alpha$ to $I\beta$. However, O-H stretching frequencies are easily affected by very small structural differences, and the Wiley-Atalla results may not indicate important differences in the hydrogen bonding schemes. It is remarkable that the structures of $I\beta$ cellulose obtained by three different methods agree so well. The unit cells varied by 0.23 Å in the x-axis direction, as well as smaller differences in the other dimensions, and the energy functions differed substantially among the PS-79, PLMR and MM3 methods. Finally, the models in the earlier studies/ were affected by fiber diffraction intensity data, known to have low reliability (French et al., 1987), while the unit cell dimensions were the only experimental data in the present calculations. The packing energies of both phases of cellulose I and of cellulose II are all very close, and the errors in the minicrystal modeling calculations are unknown. Since fiber diffraction methods depend on modeling studies for discrimination between alternative models, conclusive results will be difficult to obtain, even when diffraction intensities are available.

Conclusions

Available experimental diffraction and calorimetric data for cellulose I and II have been rationalized by computer models. Because the predicted energies were in qualitative agreement with experimental values, and because the structure predicted for cellulose $I\beta$ matched two based on fiber diffraction data, the method herein appears to be worth further development. According to these calculations, cellulose $I\alpha$ and $I\beta$ are each at local minima and thus metastable.

References

- A. Aabloo and A. D. French, Preliminary potential energy calculations of cellulose Ia crystal structure, *Macromolekul. Chem., Theory and Simulations*, in press.
- N. L. Allinger, Y. H. Yuh and J.-H. Lii, Molecular Mechanics. The MM3 Force Field for Hydrocarbons. *J. Amer. Chem. Soc.* **111**, 8551-8566 (1989)
- N. L. Allinger, M. Rahman and J.-H. Lii, A molecular mechanics force field (MM3) for alcohols and ethers. *J. Amer. Chem. Soc.* **112**, 8293-8307 (1990)
- S. Arnott and W. E. Scott, Accurate X-ray diffraction analysis of fibrous polysaccharides containing pyranose rings. Part I. The linked-atom approach, *J. Chem Soc Perkin II*, 324 (1972)
- R. H. Atalla and D. L. VanderHart, Native cellulose: a composite of two distinct crystalline forms, *Science* **223**, 283 (1984)
- A. D. French and P. S. Howley, Comparisons of structures proposed for cellulose, *Cellulose and Wood - Chemistry and Technology*, C. Schuerch, ed., John Wiley & Sons, New York, p. 159-167, 1989
- A. D. French, D. P. Miller and A. Aabloo, Miniature crystal models of cellulose polymorphs and other carbohydrates, *Int. J. Biol. Macromol.*, **15**, 1993, 30-36.
- A. D. French, W. A. Roughead and D. P. Miller, X-ray diffraction studies of ramie cellulose I. *ACS Symposium Series* **340**, 15-37 (1987)
- G. Honjo and M. Watanabe, Examination of cellulose fibers by the low-temperature specimen method of electron diffraction and electron microscopy. *Nature* **181**, 326-328 (1958)
- F. Horii, H. Yamamoto, R. Kitamura, M. Tanahashi and T. Higuchi, Transformation of native cellulose crystals induced by saturated steam at high temperatures. *Macromolecules* **20**, 2946-2949 (1987)
- J. H. Lii and N. L. Allinger, The MM3 force field for amides, polypeptides and proteins, *J. Comput. Chem.* **12**, 186-199 (1991)
- J. Mann, L. Roldan-Gonzalez and H. J. Wellard, Crystalline modifications of cellulose. Part IV. Determination of X-ray intensity data. *J. Polym. Sci.*, **47**, 165-171 (1960)
- R. - H. Mikelsaar, New space-filling atomic-molecular models, *Trends in Biotechnology* **4**, 162-163 (1986)
- H. Nishimura, T. Okano and A. Sarko, Mercerization of cellulose. 5. Crystal and molecular structure of Na-cellulose I. *Macromolecules* **24**, 759-770 (1991)
- H. Nishimura and A. Sarko, Mercerization of cellulose. 6. Crystal and molecular structure of Na-cellulose IV. *Macromolecules* **24**, 771-778 (1991)
- A. J. Pertsin, A. I. Kitaigorodsky, "The Atom-Atom Potential Method Applications to Organic Molecular Solids", Springer Ser. Chem Phys., **V43**, Springer Verlag, Berlin, 1987
- A. J. Pertsin, A. I. Kitaigorodsky and G. N. Marchenko, Crystal structure of cellulose polymorphs by potential energy calculations: 2. Regenerated and native celluloses. *Polymer* **27**, 597-601 (1986)
- A. J. Pertsin, O. K. Nugmanov, G. N. Marchenko, A. I. Kitaigorodsky, Crystal structure of cellulose polymorphs by potential energy calculations: 1. Most probable models for mercerized cellulose. *Polymer* **25**, 107-114 (1984)
- B. G. Rånby, The mercerization of cellulose. I. A thermodynamic discussion. *Acta Chem. Scand.* **6**, 101-115 (1952)
- J. Sugiyama, R. Vuong and H. Chanzy, Electron diffraction study on the two crystalline phases occurring in native cellulose from an algal cell wall. *Macromolecules* **24**, 4168-4175 (1991)
- D. L. VanderHart and R. H. Atalla, Studies of microstructure in native celluloses using solid state ¹³C NMR. *Macromolecules* **17**, 1465-1472 (1984)
- J. H. Wiley and R. H. Atalla, Raman spectra of celluloses, *ACS Symposium Series* **340**, 151-168 (1987)
- C. Woodcock and A. Sarko, Packing analysis of carbohydrates and polysaccharides. 11. Molecular and crystal structure of native ramie cellulose. *Macromolecules* **13**, 1183 (1980)
- P. Zugenmaier and A. Sarko, The variable virtual bond, *ACS Symposium Series* **141**, 226-237 (1980).

Table 1. Most probable PLMR models of cellulose Ia

U - "up" models; D - "down" models; $r_1 = \text{C4-C5-C6-O6}$; $r_2 = \text{C5-C6-O6-H}$; $r_3 = \text{C1-C2-O2-H}$; $r_4 = \text{C2-C3-O3-H}$; $r_5 = \text{C4'-C5'-C6'-O6'}$; $r_6 = \text{C5'-C6'-O6'-H'}$; $r_7 = \text{C1'-C2'-O2'-H'}$; $r_8 = \text{C2'-C3'-O3'-H'}$; $\phi_1 = \text{C1-O4'-C4'}$; $\phi_2 = \text{O5-C1-O1-C4'}$; $\phi_3 = \text{C1-O1-C4'-C5'}$; ϵ - angle describing chain rotation.

| Energy | r_1 | r_2 | r_3 | r_4 | ϕ_1 | ϕ_2 | ϕ_3 | r_5 | r_6 | r_7 | r_8 | ϵ | |
|----------------|-------|-------|-------|-------|----------|----------|----------|-------|-------|-------|-------|------------|----|
| kcal/mol | | | | | | | | | | | | | |
| U ₁ | -19.5 | -71 | -116 | 52 | -175 | 118 | -93 | -146 | -73 | -166 | 52 | -175 | 77 |
| D ₁ | -16.8 | -59 | -166 | 53 | -168 | 117 | -87 | -143 | -74 | -163 | 53 | -170 | 1 |
| U ₂ | -16.3 | -69 | 61 | 164 | -168 | 119 | -92 | -148 | -72 | 63 | 170 | -171 | 72 |
| U ₃ | -16.2 | -68 | -102 | 56 | -161 | 115 | -94 | -147 | -69 | -102 | 56 | -161 | 78 |
| U ₄ | -15.7 | -64 | -60 | 79 | -177 | 116 | -95 | -146 | -66 | -57 | 79 | -177 | 79 |
| U ₅ | -15.0 | -69 | -34 | 54 | -72 | 116 | -94 | -146 | -71 | -35 | 54 | -72 | 78 |

Table 2. Most probable PLMR models of cellulose Ib

Variables: $r_{11}, r_{12} = \text{C4-C5-C6-O6}$ for a corner and a central chain; $r_{21}, r_{22} = \text{C5-C6-O6-H}$ for a corner and a central chain; $r_{31}, r_{32} = \text{C1-C2-O2-H}$ for a corner and a central chain; $r_{41}, r_{42} = \text{C2-C3-O3-H}$ for a corner and a central chain; $\varphi_1, \varphi_2 =$ angles describing the chain rotation; $\delta_1, \delta_2 =$ angles describing a virtual bond rotation; $s =$ a vertical shift of a central chain (unit = 1/10c).

| E | r_{11} | r_{21} | r_{31} | r_{41} | r_{12} | r_{22} | r_{32} | r_{42} | δ_1 | δ_2 | φ_1 | φ_2 | s | |
|----------------|----------|----------|----------|----------|----------|----------|----------|----------|------------|------------|-------------|-------------|------|------|
| kcal/mol | | | | | | | | | | | | | | |
| U ₁ | -19.9 | -71 | -166 | 52 | -175 | -71 | -166 | 52 | -175 | 42 | 42 | -127 | -126 | -2.7 |
| U ₂ | -19.0 | -67 | -160 | 61 | 162 | -68 | -160 | 60 | 161 | 44 | 44 | -122 | -128 | 2.6 |
| U ₃ | -18.7 | -68 | -67 | 62 | -178 | -70 | -63 | 60 | -176 | 44 | 44 | -118 | -135 | 3.4 |
| U ₄ | -17.8 | -67 | -64 | 61 | 163 | -69 | -62 | 60 | 163 | 43 | 43 | -116 | -136 | 3.3 |
| U ₅ | -16.6 | -63 | -56 | 61 | -59 | -68 | -61 | 60 | -61 | 46 | 43 | -120 | -130 | 3.1 |
| U ₆ | -16.4 | -72 | 60 | -160 | 180 | -72 | 67 | -149 | -176 | 41 | 42 | -125 | -124 | 2.6 |
| U ₇ | -16.1 | -72 | 58 | -172 | -61 | -72 | 60 | -167 | -65 | 40 | 42 | -125 | -124 | 2.7 |

Table 3. Most probable models of cellulose Ia and Ib.

U - "up" models; D - "down" models; α - models of phase Ia; β - models of phase Ib.

| Model | | Energy kcal/mol | | Mean atom movement | | H-bonds in chain | H-bonds in sheet | H-bonds found by MM3 |
|----------------|----|-----------------|-----|--------------------|-------------|--------------------|------------------|----------------------|
| | | PLMR | MM3 | with H's | without H's | | | |
| U ₁ | tg | -19.5 | 185 | .175 | .106 | O5..HO3 O6..HO2 | O3..HO6 | |
| D ₁ | tg | -16.8 | 206 | .158 | .108 | O5..HO3 O6..HO2 | O3..HO6 | |
| U ₂ | tg | -16.3 | 222 | .178 | .148 | O5..HO3 O2;O2..HO6 | | |
| U ₃ | tg | -16.2 | 218 | .177 | .092 | O5..HO3 O6..HO2 | | O3..HO6 |
| U ₄ | tg | -15.7 | 215 | .193 | .098 | O5..HO3 O6..HO2 | | O3..HO6 |
| U ₅ | tg | -15.0 | 234 | .056 | .035 | O6..HO2 | O6..HO3 | |
| U ₁ | tg | -19.9 | 182 | .187 | .167 | O6..HO3 O6..HO2 | O3..HO6 | |
| U ₂ | tg | -19.0 | 185 | .186 | .153 | O5..HO3 O6..HO2 | O3..HO6 | |
| U ₃ | tg | -18.7 | 200 | .176 | .140 | O5..HO3 O6..HO2 | | |
| U ₄ | tg | -17.8 | 196 | .185 | .143 | O5..HO3 O6..HO2 | | |
| U ₅ | tg | -16.6 | 212 | .201 | .177 | O6..HO2 | O6..HO3 | |
| U ₆ | tg | -16.4 | 200 | .267 | .234 | O5..HO3 O2;O4..HO6 | | O6..HO2 |
| U ₇ | tg | -16.1 | 207 | .254 | .199 | O2;O4..HO6 | O6..HO3 | |

Table 4. Cartesian coordinates (Å) of the best model (U₀1) of the cellulose Ia crystal.

| Name | X | Y | Z | Name | X | Y | Z |
|--------|--------|--------|-------|---------|--------|--------|-------|
| 1 O2 | .183 | -.809 | .000 | 22 O2' | -1.708 | 1.162 | -.281 |
| 2 O3 | .300 | -.072 | 1.171 | 23 O3' | -.288 | .065 | 6.338 |
| 3 O4 | .860 | -.976 | 2.261 | 24 O4' | -.888 | .949 | 7.424 |
| 4 O5 | .928 | -.267 | 3.508 | 25 O5' | -.945 | .235 | 8.668 |
| 5 O6 | -.364 | .141 | 3.965 | 26 O6' | .354 | -.133 | 9.137 |
| 6 C1 | -.968 | 1.114 | 2.961 | 27 C1' | 1.000 | -1.083 | 8.138 |
| 7 C2 | -1.049 | .475 | 1.581 | 28 C2' | 1.074 | -.439 | 6.760 |
| 8 C3 | 2.261 | -1.462 | 1.955 | 29 C3' | -2.300 | 1.390 | 7.104 |
| 9 C4 | -.183 | .809 | 5.169 | 30 C4' | 2.301 | -1.441 | 8.600 |
| 10 C5 | -2.262 | 1.512 | 3.411 | 31 C5' | 1.595 | -1.376 | 5.820 |
| 11 C6 | -1.530 | 1.431 | .639 | 32 C6' | -.933 | -.784 | 6.167 |
| 12 H1 | .973 | .757 | 1.010 | 33 H1' | -.281 | 1.835 | 7.544 |
| 13 H2 | .224 | -1.842 | 2.373 | 34 H2' | .964 | .752 | 9.239 |
| 14 H3 | -1.003 | -.724 | 4.059 | 35 H3' | .421 | -1.993 | 8.080 |
| 15 H4 | -.360 | 2.005 | 2.911 | 36 H4' | 1.718 | .427 | 6.801 |
| 16 H5 | -1.721 | -.370 | 1.613 | 37 H5' | -2.326 | 2.379 | 6.077 |
| 17 H61 | 2.272 | -2.408 | .888 | 38 H61' | -2.884 | .535 | 6.796 |
| 18 H62 | 2.886 | -.620 | 1.696 | 39 H62' | -2.765 | 1.784 | 7.996 |
| 19 HO2 | 2.688 | -1.911 | 2.839 | 40 HO2' | -3.137 | 2.898 | 5.921 |
| 20 HO3 | 3.065 | -2.952 | .726 | 41 HO3' | 2.419 | -1.786 | 9.504 |
| 21 HO6 | -2.378 | 1.860 | 4.315 | 42 HO6' | 1.770 | -1.100 | 4.902 |

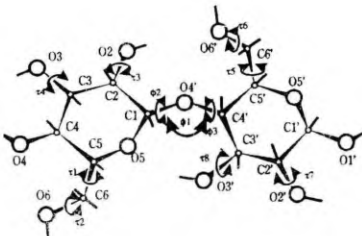


Figure 1. The variable angles of cellulose unit PLMR

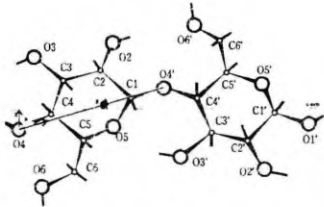


Figure 2. A virtual bond constructed for a unit of cellulose

Table 5. Carthesian coordinates (Å)of the best model (U_β1) of cellulose I_β crystal.

| Name | x | y | z | Name | x | y | z |
|--------------|--------|--------|-------|---------------|-------|-------|--------|
| Corner Chain | | | | Central chain | | | |
| 1 O2 | -.339 | 2.694 | 3.434 | 1 O2 | 3.573 | 6.262 | .685 |
| 2 O3 | .084 | 2.107 | .658 | 2 O3 | 4.009 | 5.684 | -2.091 |
| 3 O4 | -.487 | -.651 | .000 | 3 O4 | 3.500 | 2.915 | -2.750 |
| 4 O5 | .403 | -.882 | 3.508 | 4 O5 | 4.396 | 2.703 | .759 |
| 5 O6 | -.326 | -3.280 | .867 | 5 O6 | 3.718 | .289 | -1.882 |
| 6 C1 | -.134 | .358 | 3.972 | 6 C1 | 3.831 | 3.931 | 1.223 |
| 7 C2 | .208 | 1.459 | 2.977 | 7 C2 | 4.147 | 5.039 | .227 |
| 8 C3 | -.325 | 1.111 | 1.593 | 8 C3 | 3.622 | 4.679 | -1.156 |
| 9 C4 | .145 | -.264 | 1.175 | 9 C4 | 4.123 | 3.315 | -1.574 |
| 10 C5 | -.174 | -1.287 | 2.257 | 10 C5 | 3.827 | 2.286 | -.492 |
| 11 C6 | .376 | -2.662 | 1.943 | 11 C6 | 4.408 | .923 | -.807 |
| 12 H1 | -1.207 | .274 | 4.063 | 12 H1 | 2.760 | 3.823 | 1.314 |
| 13 H2 | 1.279 | 1.582 | 2.930 | 13 H2 | 5.216 | 5.187 | .180 |
| 14 H3 | -1.404 | 1.066 | 1.623 | 14 H3 | 2.545 | 4.610 | -1.126 |
| 15 H4 | 1.213 | -.230 | 1.016 | 15 H4 | 5.190 | 3.374 | -1.734 |
| 16 H5 | -1.245 | -1.374 | 2.366 | 16 H5 | 2.758 | 2.175 | -.383 |
| 17 H61 | 1.423 | -2.582 | 1.691 | 17 H61 | 5.453 | 1.027 | -1.060 |
| 18 H62 | .310 | -3.286 | 2.822 | 18 H62 | 4.358 | .298 | .072 |
| 19 HO2 | -.152 | 3.005 | 4.339 | 19 HO2 | 3.753 | 6.576 | 1.590 |
| 20 HO3 | -.233 | 2.071 | -.263 | 20 HO3 | 3.693 | 5.641 | -3.012 |
| 21 HO6 | -.219 | -4.234 | .699 | 21 HO6 | 3.847 | -.663 | -2.050 |
| 22 O2' | .339 | -2.694 | 8.614 | 22 O2' | 4.372 | .891 | 5.865 |
| 23 O3' | -.084 | -2.107 | 5.838 | 23 O3' | 3.936 | 1.468 | 3.089 |
| 24 O4' | .487 | .651 | 5.180 | 24 O4' | 4.445 | 4.237 | 2.430 |
| 25 O5' | -.403 | .882 | 8.688 | 25 O5' | 3.549 | 4.449 | 5.939 |
| 26 O6' | .326 | 3.280 | 6.047 | 26 O6' | 4.227 | 6.863 | 3.298 |
| 27 C1' | .134 | -.358 | 9.152 | 27 C1' | 4.114 | 3.221 | 6.403 |
| 28 C2' | -.208 | -1.459 | 8.157 | 28 C2' | 3.798 | 2.113 | 5.407 |
| 29 C3' | .325 | -1.111 | 6.773 | 29 C3' | 4.323 | 2.473 | 4.024 |
| 30 C4' | -.145 | .264 | 6.355 | 30 C4' | 3.822 | 3.837 | 3.606 |
| 31 C5' | .174 | 1.287 | 7.437 | 31 C5' | 4.118 | 4.867 | 4.688 |
| 32 C6' | -.376 | 2.662 | 7.123 | 32 C6' | 3.537 | 6.229 | 4.373 |
| 33 H1' | 1.207 | -.274 | 9.243 | 33 H1' | 5.185 | 3.329 | 6.494 |
| 34 H2' | -1.279 | -1.582 | 8.110 | 34 H2' | 2.729 | 1.966 | 5.360 |
| 35 H3' | 1.404 | -1.066 | 6.803 | 35 H3' | 5.400 | 2.542 | 4.054 |
| 36 H4' | -1.213 | .230 | 6.196 | 36 H4' | 2.755 | 3.778 | 3.446 |
| 37 H5' | 1.245 | 1.374 | 7.546 | 37 H5' | 5.187 | 4.978 | 4.797 |
| 38 H61' | -1.423 | 2.582 | 6.871 | 38 H61' | 2.492 | 6.125 | 4.120 |
| 39 H62' | -.310 | 3.286 | 8.002 | 39 H62' | 3.587 | 6.854 | 5.252 |
| 40 HO2' | .152 | -3.005 | 9.519 | 40 HO2' | 4.192 | .576 | 6.770 |
| 41 HO3' | .233 | -2.071 | 4.917 | 41 HO3' | 4.252 | 1.511 | 2.168 |
| 42 HO6' | .219 | 4.234 | 5.879 | 42 HO6' | 4.098 | 7.815 | 3.130 |

PARALLEL AND ANTIPARALLEL MODELS FOR CRYSTALLINE PHASES OF NATIVE CELLULOSE

Parallel and antiparallel models of cellulose

Raik-Hiio Mikelsaar* Institute of General and Molecular Pathology, Tartu University, Veski Street 34, Tartu EE2400, Estonia; Alvo Aabloo Institute of Experimental Physics and Technology, Tartu University, Tähe Street 4, Tartu EE2400, Estonia.

ABSTRACT

Tartu plastic space-filling atomic-molecular models were used to investigate the structure of cellulose (*I α* and *I β*) phases. It was elucidated that both parallel and antiparallel structures can be accommodated with unit cell geometries described by Sugiyama *et al.* (1991). Packing energy calculations revealed that parallel models of *I α* and *I β* phases have very close energy. In contrast, in case of antiparallel structure models, *I β* is energetically much more favourable in comparison with *I α* phase. Our data indicate that antiparallel structure proposed for cellulose *I α* phase is metastable and should be easily converted to *I β* phase. So the antiparallel models are more suitable for explanation of cellulose *I α* →*I β* →II interconversion.

Keywords: cellulose molecular and crystalline structure, molecular modelling by plastic atomic models, packing energy calculations.

INTRODUCTION

Native cellulose I is a wide-spread biopolymer which, through either mercerization or regeneration, can be converted irreversibly into another form - cellulose II. It is generally accepted that both mercerized and regenerated cellulose II have nearly identical crystal structures with two-chain unit cells (Stipanovic and Sarko, 1976; Kolpak and Blackwell, 1976; Kolpak *et al.*, 1978; Pertsin *et al.*, 1984; Nishimura *et al.*, 1991; Nishimura and Sarko, 1991). In contrast, IR-spectroscopy, X-ray, electron diffraction and NMR methods revealed that cellulose I occurs in two forms: type IA (algal-bacterial) cellulose and type IB (ramie-cotton) cellulose (Marrinan and Mann, 1956; Honjo and Watanabe, 1958; Fisher and Mann, 1960; Hebert, 1985; Horii *et al.*, 1987a). On 1984 by NMR investigations it was elucidated that both forms are a mixture of two crystalline phases (*α* and *β*) which proportion depends on the source of cellulose (Atalla and VanderHart, 1984; VanderHart and Atalla, 1984). Also with NMR spectroscopy Horii *et al.* (1987b) showed that the *I α* phase of cellulose is metastable and, through a hydrothermal annealing treatment, can be converted readily into the thermodynamically stable *I β* phase. Recently Sugiyama *et al.* (1991) studied cellulose from a

green alga *Microdictyon* by electron diffraction. It was found that the major *I α* phase has a one-chain, triclinic (P1) structure and the minor *I β* phase is characterized by two-chain unit cell and a monoclinic (P2₁) structure. This study disclosed the unit cell parameters but did not present a detailed molecular structure of these native cellulose crystalline phases.

The aim of the current paper was using molecular modelling method and packing energy calculations to elucidate the most favourable stereochemical structures corresponding to unit cell parameters found by Sugiyama *et al.* (1991) in cellulose *I α* and *I β* crystalline phases.

MATERIAL AND METHODS

Molecular modelling was carried out by Tartu plastic space-filling models having improved parameters and design (Mikelsaar *et al.*, 1985; Mikelsaar 1986). *Packing energy calculations* were carried out by the PLMR program, a rigid-ring method (Pertsin *et al.*, 1984).

To present *graphically* the stereochemical models investigated, we drew unit cell projections on the plane crossing the c-axis at 90°. Cross-sections of *up* cellulose chains were given by white and *down* chains by grey ovals. The number of c/4 shifts of molecules in the direction of the c-axis is noted by numbers inside of the ovals.

To make a short *digital* designation of the cellulose models: 1) symbols "P" and "A" were used to represent correspondingly "parallel" and "antiparallel", 2) numbers indicating c/4 shifts were applied for characterizing of each chain in the unit cell, 3) underlining was used to differentiate *down* chains from *up* chains.

RESULTS

1. *Molecular models fitting with cellulose I α unit cell parameters*

1.1. *Model P1*

Fig. 1

The investigation of green alga *Microdictyon* by Sugiyama *et al.* (1991) is the first work where unit cell parameters are given for pure cellulose *I α* microcrystals: $a = 0.674$ nm, $b = 0.593$ nm, $c = 1.036$ nm, $\alpha = 117^\circ$, $\beta = 113^\circ$ and $\gamma = 81^\circ$. One-chain unit cells result in parallel packing of molecular chains. In Figure 1 there are our graphic presentations of above-described one-chain unit cell and sterically corresponding two- and eight-chain cells on the plane crossing the c-axis at 90°.

Molecular modelling by Tartu devices showed that if there are the usual intramolecular bonds O3-H...O5 and O2-H...O6 the hydroxymethyl groups should be in *tg* conformation and all parallel chains are bound by intrasheet bonds O6-H...O3 (Fig. 2 and Fig.3).

Fig. 2

Fig. 3

Packing energy calculations of the above-described cellulose structure imitation revealed that model P1 has potential energy -19.5 kcal/mol.

1.2. Models A1a and A1b

Fig. 4

Fig. 5

Two antiparallel models A1a and A1b characterized by eight-chain unit cell may be geometrically derived from the one-chain unit cell described by Sugiyama *et al.* (1991) (Fig. 4 and 5). The geometry of these two models is very similar. The only difference is that in A1a parallel chains follow each other from down right to up left dividing major angles of unit cell and in A1b from down left to up right dividing minor angles of unit cell. So A1a may be called "left" and A1b "right" type of structures.

Fig. 6

Fig. 7

According to molecular modelling data, chains having usual intramolecular H-bonds and *tg* conformation of hydroxymethyl groups should be in models A1a and A1b bound by intrasheet bonds O6-H...O2 (Fig. 6 and 7).

The potential energy values of above-described structures are rather similar: model A1a has - 16.4 kcal/mol and A1b - 15.5 kcal/mol.

2. Molecular models fitting with cellulose I β unit cell parameters

2.1. Model P2

Sugiyama *et al.* (1991) have established that I β phase of Microdictyon cellulose has a two-chain monoclinic structure with unit cell parameters of $a = 0.801$ nm, $b = 0.817$ nm, $c = 1.036$ nm and $\gamma = 97.3^\circ$. The authors suppose that this structure consists of parallel molecules and the center and corner chains are staggered by $c/4$. Figure 8 shows our graphic presentation of above-mentioned structure.

Fig. 8

Molecular modelling revealed that intra- and intermolecular H-bonds in model P2 are identical to those in model P1 (Fig. 2 and 3).

The energy value of model P2 is close to the model P1 one: -19.9 kcal/mol.

2.2. Model A2

A graphic presentation of the structure corresponding to model A2 is demonstrated in Figure 9.

Fig. 9

Molecular modelling showed that chain conformation and intrasheet H-bonding network in model A2 are identical to the model P2 one's. The only difference is that sheets are parallel in model P2 and antiparallel in model A2.

Potential energy of model A2 is relatively high: - 15.8 kcal/mol.

2.3. Models A3a and A3b

Fig. 10

Fig. 11

Two antiparallel models A3a and A3b characterized by eight-chain unit cell may be geometrically derived from the two-chain unit cell described by Sugiyama *et al.* (1991) (Fig. 10 and 11). These structures are similar to the "left" and "right" models A1a and A1b differing mainly by the presence of intrasheet stagger of cellulose chains.

Fig. 12

Fig. 13

The investigation by plastic models indicated that in case of the preservation of standard intramolecular H-bonds and *tg* conformation of hydroxymethyl groups the antiparallel molecular chains are bound by bonds O6-H...O3 (Fig. 12 and 13).

There is a considerable difference between energies of above-described two structures: the "left" type model A3a has -21.0 kcal/mol and "right" model A3b -15.1 kcal/mol.

Table 1

A generalized characterization of above-described cellulose models is presented in Table 1.

DISCUSSION

Our investigation revealed that both parallel and antiparallel structure models can be proposed, based on cellulose *I α* and *I β* crystalline phase unit cell parameters published by Sugiyama *et al.* (1991). Although all they have usual chain conformation and intramolecular H-bonding network, the potential energy values between seemingly close structures in certain cases are very different.

Sugiyama *et al.* (1991) interpret their experimental data in the light of all-parallel-structure view. Following this view, the *I α* phase structure should correspond to our model P1 and *I β* phase to model P2. However, the conception of Sugiyama *et al.* (1991) meets many difficulties when one try to explain the molecular mechanisms of cellulose *I α* →*I β* →II interconversion.

First, if the structures of *I α* and *I β* phases correspond to the models P1 and P2, the only change after the process of *I α* →*I β* conversion will be a *c*/2 shift of each third and fourth molecular sheet. The intra- and intersheet distances between molecular chains should remain

unchanged. However the comparison of $I\beta$ phase two-chain unit cell parameters in the work of Sugiyama *et al.* (1991) with corresponding data derived by us from $I\alpha$ one-chain unit cell indicates that there are rather considerable differences: 1) unit cell of $I\beta$ phase: $a = 0.801$ nm, $b = 0.817$ nm, $\gamma = 97.3^\circ$, and 2) corresponding unit cell derived from $I\alpha$ phase: $a = 0.814$ nm, $b = 0.825$ nm, $\gamma = 99.2^\circ$.

Second, if the only rearrangement in the $I\alpha \rightarrow I\beta$ conversion was a $c/2$ shift of chains, the H-bonding network in the cellulose structure must remain unchanged. However, Raman spectroscopic data reveal that H-bonds in $I\alpha$ phase cellulose are different from those of $I\beta$ (Atalla, 1989).

Third, if there only $c/2$ shift of chains occurs and sheets are bound only by weak VanderWaals' interactions, the energy barrier between molecular structures of $I\alpha$ and $I\beta$ phases should be very small. It would be expected that cellulose I crystallites contain equal amounts of both phases. However, $I\alpha$ dominates in IA and $I\beta$ in IB celluloses.

Fourth, because annealing of cellulose $I\alpha$ converts it irreversibly to $I\beta$, the energy of stable $I\beta$ should be considerably lower than the metastable $I\alpha$ energy. However, the packing energy difference between P1 and P2 models is very small.

Fifth, cellulose II is generally considered to have antiparallel molecular structure. Sarko *et al.* (1987) and Nishimura *et al.* (1991) established that already Na-cellulose I, formed during the initial step of a controlled mercerization of ramie cellulose, has antiparallel-chain structure. Because the change of chain direction in this phase is not possible, it would be very difficult to imagine how an antiparallel structure could arise from the parallel one. It is true that antiparallel structure of cellulose II can arise through a process of "interdigitation" (see French, 1985) and energy minimization procedure reveal a possibility of parallel chain packing in cellulose II (Sakthivel *et al.*, 1988), but these arguments seem to be not very convincing. In our opinion the molecular mechanisms of cellulose $I\alpha \rightarrow I\beta \rightarrow II$ interconversion can be easier explained in the light of antiparallel structure models.

Comparing cellulose parallel model P1 with antiparallel models A1a and A1b one can see that these models are sterically very similar. If the H-bonding network and chain polarity are not clearly established, the structure characterized by eight-chain unit cell can be considered as structure of one-chain unit cell. We think that antiparallel model A1a having lower energy (-16.4 kcal/mol) in comparison with the model A1b (-15.5 kcal/mol) is a possible pretender for a "masked" eight-chain unit cell in the cellulose $I\alpha$ crystal structure described by Sugiyama *et al.* (1991). It may be that the experimental procedure of electron diffraction in above-mentioned investigation allowed to record only the parameters of subunits of eight-chain unit cells. The antiparallel structure is not excluded also for cellulose $I\beta$ crystalline phase because these authors marked that the exact position of the cellulose chains in the monoclinic unit cell was even more difficult to ascertain as long as the structure refinement was not achieved. No doubt, A3a is the best antiparallel model for fitting to the unit cell structure of $I\beta$

crystalline phase because it has an excellent packing energy: -21.0 kcal/mol. For both A1 and A3 models, the "left type" structures A1a and A3a have lower energy than the "right type" A1b and A3b models.

If the "left type" antiparallel models A1a and A3a are thought to be the candidates for structures of cellulose I α and I β crystalline phases, correspondingly, it would be easy to explain the experimental data on the interconversion of these phases. Model A1a, having relatively high energy -16.4 kcal/mol, is metastable and can during annealing process easily to be converted irreversibly into the model A3a, having very low energy -21.0 kcal/mol. Because the models A1a and A3a have different H-bonding network, our conception on antiparallel cellulose structure is in accordance also with Raman-spectroscopic data of Atalla (1989). The high energy barrier between I α and I β crystalline phases, bound with rearrangement of hydrogen bonds, may be the reason why the proportion of these phases in different samples of native cellulose is relatively constant. The notable structural differences between models A1a and A3a allow one also to understand the differences between corresponding unit cell parameters of I α and I β cellulose crystalline phases.

The order of chain rearrangements in the transformation of models A1a \rightarrow A3a, probably corresponding to the cellulose I α \rightarrow I β phase conversion, seems to be simple. After the interruption in model A1a of intrasheet bonds O2-H...O6 the alternate sheets of parallel chains must move one step diagonally in the plane of these sheets and new intrasheet bonds O3-H...O6 characteristic of model A3a should be formed.

The process of cellulose I β \rightarrow II conversion can be also easily interpreted by all-antiparallel conception. Because already Na-cellulose I, formed in the initial step of mercerization, has antiparallel structure (Sarko *et al.*, 1987; Nishimura *et al.*, 1991), no change of chain direction is needed. The conversion from the structure corresponding to model A3a to Na-cellulose I may contain following steps: 1) interruption of intrasheet H-bonds, 2) including of Na⁺ and OH⁻ ions and water molecules, 3) forming of a large four-chain unit cell. The intersheet hydrogen bonding found in cellulose II is one of the most significant differences between this structure and the native cellulose and may be the reason of irreversibility of cellulose I \rightarrow II conversion (Kolpak *et al.*, 1978)

In some electron-microscopical studies it was shown that the reducing ends of cellulose exposed at the tip of a fragmented microfibril of Valonia and Acetobacter can be asymmetrically labeled with silver indicating a unidirectional alignment of glucan chains (Hieta *et al.*, 1984; Kuga and Brown, 1988). The same feature was visualized also by Chanzy and Henrissat (1985), who showed that the digestion of Valonia cellulose by cellobiohydrolase proceeds in one direction in a single microfibril. These data have been used in favour of conception on parallel cellulose structure. In our opinion these results may equally well be explained from the position of antiparallel conception. For example, according to our model A1a it may be supposed that crystallites of cellulose I α phase contain on one surface a layer

including exclusively parallel molecular chains. Probably the asymmetrical silver labeling and enzymatic digestion will occur on this layer.

Summing up, molecular modelling by plastic models and packing energy calculations revealed that both parallel and antiparallel structure models can be accommodated with cellulose unit cell geometries described by Sugiyama *et al.* (1991). However, the all-antiparallel conception has several advantages in comparison with all-parallel view in explaining the experimental data on molecular mechanisms of cellulose $\alpha \rightarrow \beta$ -H conversion.

REFERENCES

- Atalla, R.H.; VanderHart, D.L. Native cellulose: a composite of two distinct crystalline forms. *Science* 1984, **223**, 283-285.
- Atalla, R.H. Patterns of aggregation in native celluloses: implication of recent spectroscopic studies. *Cellulose structural and functional aspects*. Eds. J.F.Kennedy, G.O.Phillips, P.A.Williams. Ellis Horwood Ltd., New York, Chichester, Brisbane, Toronto, 1989, 61-73.
- Chanzy, H.; Henrissat, B. Unidirectional degradation of Valonia cellulose microcrystals subjected to cellulose *FEBS Lett.* 1985, **184**, 285-288.
- Fisher, D.G.; Mann, J. Crystalline modifications of cellulose. IV. Unit cell and molecular symmetry of cellulose I. *J. Polymer Sci.* 1960, **17**, 189-194.
- French, A.D. Physical and theoretical methods for determining the supramolecular structure of cellulose. In: *Cellulose chemistry and its applications*. Eds. T.P.Nevell, S.H.Zeronian 1985, 84-111.
- Hebert, J.J. Crystalline form of native celluloses. *Science* 1985, **227**, 79.
- Hieta, K.; Kuga, S.; Usuda, M. Electron staining of reducing ends evidences a parallel-chain structure in Valonia cellulose. *Biopolymers* 1984, **23**, 1807-1810.
- Honjo, G.; Watanabe, M. Examination of cellulose fibre by the low-temperature specimen method of electron diffraction and electron microscopy. *Nature* 1958, **181**, 326-328.
- Horii, F.; Hirai, A.; Kitamaru, R. CP/MAS¹³C NMR spectra of the crystalline components of native celluloses. *Macromolecules* 1987a, **20**, 2117-2120.
- Horii, F.; Yamamoto, H.; Kitamaru, R.; Tanahashi, M.; Higuchi, T. CP/MAS¹³C NMR spectra of crystalline components of native celluloses. *Macromolecules* 1987b, **20**, 2946-2949.
- Kolpak, F.J.; Blackwell J. Determination of the structure of cellulose II. *Macromolecules* 1976, **9**, 273-278.
- Kolpak, F.J.; Weih, M.; Blackwell J. Mercerization of cellulose. 1. Determination of the structure of mercerized cotton. *Polymer* 1978, **19**, 123-131.
- Kuga, S.; Brown, R.M., Jr. Silver labeling of the reducing ends of bacterial cellulose. *Carbohydrate Research* 1988, **180**, 345-350.

- Marrinan, H.J.; Mann, J. Infrared spectra of the crystalline modifications of cellulose. *J. Polymer Sci.* 1956, **21**, 301-311.
- Mikelsaar, R.-H.N.; Bruskov, V.I.; Poltev, V.I. *New precision space-filling atomic-molecular models*: Scientific Center of Biological Research of the Academy of Sciences of the USSR in Pushchino, 1985.
- Mikelsaar, R. New space-filling atomic-molecular models. *Trends Biotechnol.* 1986, **6**, 162-163.
- Nishimura, H.; Okano, T.; Sarko, A. Mercerization of cellulose. 5. Crystal and molecular structure of Na-cellulose I. *Macromolecules* 1991, **24**, 759-770.
- Nishimura, H.; Sarko, A. Mercerization of cellulose. 6. Crystal and molecular structure of Na-cellulose IV. *Macromolecules* 1991, **24**, 771-778.
- Pertsin, A.J.; Nugmanov, O.K.; Marchenko, G.N.; Kitaigorodsky, A.I. Crystal structure of cellulose polymorphs by potential energy calculations. 1. Most probable models for mercerized cellulose. *Polymer* 1984, **25**, 107-114.
- Sakthivel, A.; Turbak, A.F.; Young, R.A. New structural models for cellulose II derived from packing energy minimization. In: *Cellulose and wood - chemistry and technology*. Ed. C.Schuerch. State University of New York, College of Environmental Science and Forestry, Syracuse, New York 13210, USA, 1988, 67-76.
- Sarko, A.; Nishimura, H.; Okano, T. Crystalline alkali-cellulose complexes as intermediates during mercerization. In: *The structure of cellulose*; ACS Symposium Series 340; American Chemical Society: Washington, DC, 1987, 169-177.
- Stipanovic, A.J.; Sarko, A. Packing analysis of carbohydrates and polysaccharides. 6. Molecular and crystal structure of regenerated cellulose II. *Macromolecules* 1976, **9**, 851-857.
- Sugiyama, J.; Vuong, R.; Chanzy, H. Electron diffraction study on the two crystalline phases occurring in native cellulose from an algal cell wall. *Macromolecules* 1991, **24**, 4168-4175.
- VanderHart, D.L.; Atalla, R.H. Studies of microstructure in native cellulose using solid-state ^{13}C NMR. *Macromolecules* 1984, **17**, 1465-1472.

References

1. Sugiyama, J.; Vuong, R.; Chanzy, H. *Macromolecules* 1991, v. 24, p. 4168.
2. Atalla, R. H.; VanderHart, D. L. *Science* 1984, v. 223, p. 283.
3. VanderHart, D. L.; Atalla, R. H. *Macromolecules* 1984, v. 17, p. 1465.
4. Horii, F.; Yamamoto, H.; Kitamaru, R.; Takahashi, M.; Higuchi, T. *Macromolecules* 1987, v. 20, p. 2946.
5. Mikelsaar, R. *Trends Biotechnol.* 1986, v. 6, p. 162.
6. Pertsin, A. J.; Nugmanov, O. K.; Marchenko, G. N.; Kitaigorodsky, A. I. *Polymer* 1984, v. 25, p. 107.
7. Mikelsaar, R.-H.; Aabloo, A. In: *Cellulosics: Chemical, Biochemical and Material Aspects*. Eds. J. F. Kennedy, G. O. Phillips, P. A. Williams. Ellis Horwood, Chichester, 1993, p. 57.
8. Atalla, R. H. In: *Cellulose Structural and Functional Aspects*. Eds. J. F. Kennedy, G. O. Phillips, P. A. Williams. Ellis Horwood, Chichester, 1989, p. 61.
9. Sarko, A.; Nishimura, H.; Okano, T. In: *Structure of Cellulose*; ACS Symposium Series 340; American Chemical Society: Washington, DC, 1987 p. 169.
10. Nishimura, H.; Okano, T.; Sarko, A. *Macromolecules* 1991, v. 24, p. 759.

Table 1.

Parallel and antiparallel stereochemical models for cellulose $I\alpha$ and $I\beta$ crystalline phases.

| Phase | Model | Unit cell designation | Intramolecular H-bonds | Intrasheet H-bonds | -CH ₂ OH group conformation | Packing energy kcal/mol |
|-----------|-------|-----------------------|------------------------|--------------------|--|-------------------------|
| $I\alpha$ | P1 | 1 | O3-H...O5 O2-H...O6 | O6-H...O3 | tg | -19.5 |
| $I\beta$ | P2 | O1 | O3-H...O5 O2-H...O6 | O6-H...O3 | tg | -19.9 |
| $I\alpha$ | A1a | 00112233 | O3-H...O5 O2-H...O6 | O6-H...O2 | tg | -16.4 |
| $I\alpha$ | A1b | 00112233 | O3-H...O5 O2-H...O6 | O6-H...O2 | tg | -15.5 |
| $I\beta$ | A2 | 01 | O3-H...O5 O2-H...O6 | O6-H...O3 | tg | -15.8 |
| $I\beta$ | A3a | 01123243 | O3-H...O5 O2-H...O6 | O6-H...O3 | tg | -21.0 |
| $I\beta$ | A3b | 01213234 | O3-H...O5 O2-H...O6 | O6-H...O3 | tg | -15.1 |

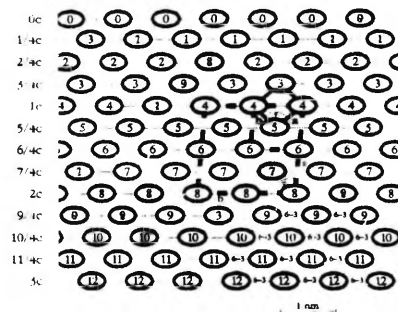


Fig. 1 Model P1 One-chain unit cell: 1. The projection of one-chain unit cell described by Sugiyama et al (1991) for cellulose IX phase is compared with projections of sterically corresponding two-chain and eight-chain unit cells on the plane crossing c-axis at 90°. Symbols: white ovals - cross sections of up chains; number in ovals - the amount of c/4 shifts in the direction of c-axis: a, b, c, γ - unit cell parameters; 6-3 intrasheet bonds O5-H...O3.

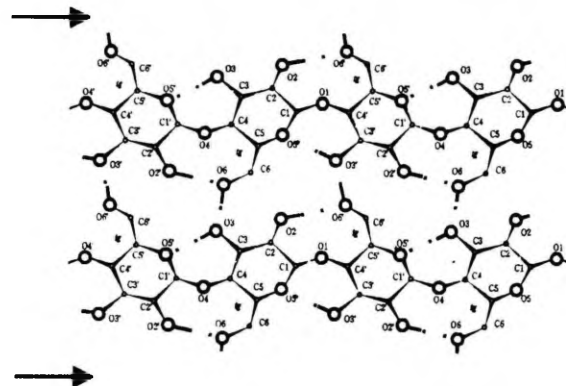


Fig. 2 Model P1. Stereochemical formula with H-bonding network.

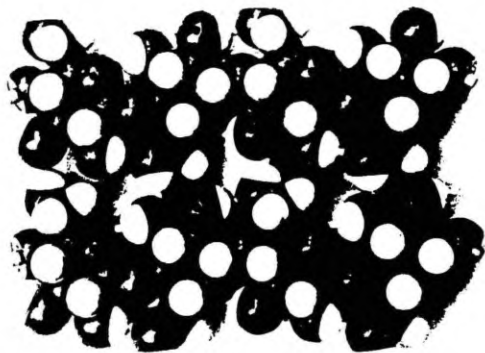


Fig. 3 Model P1 Plastic space-filling molecular model - Cf. Fig. 2

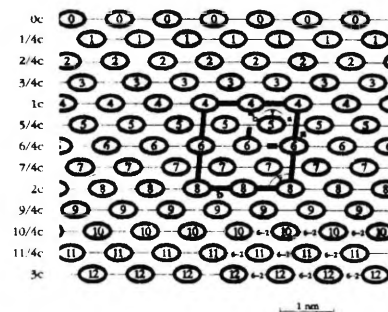


Fig. 4 Model A1a. Eight-chain unit cell: 00112233. Symbols: grey ovals - cross sections of down chains; 6-2 - intrasheet bonds O6-H...O2; other symbols see Fig. 1.

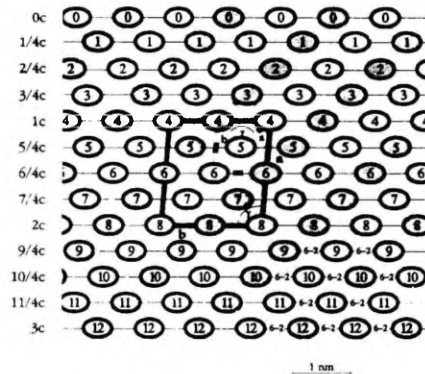


Fig. 5. Model A1b. Eight-chain unit cell. 00112233. Symbols: see Fig. 1 and 4.

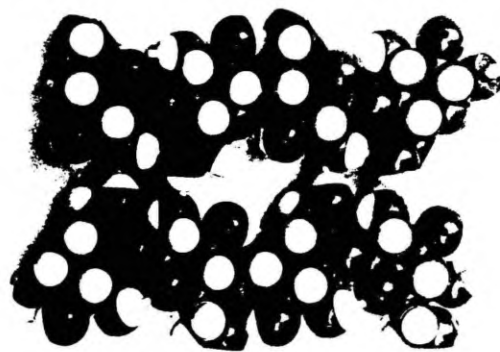


Fig. 7. Models A1a and A1b. Plastic space-filling molecular model. Cf. Fig. 6.

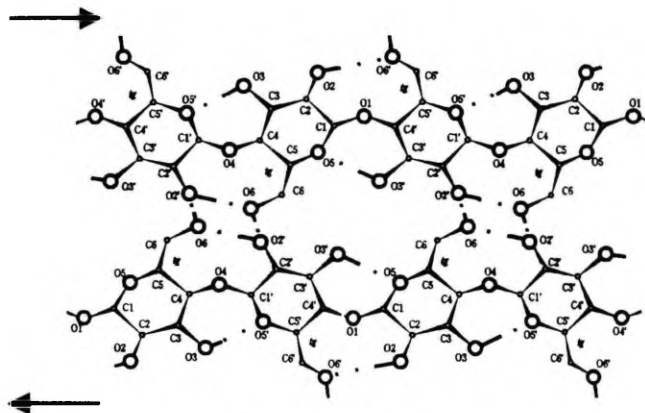


Fig. 6. Models A1a and A1b. Stereochemical formula with H-bonding network.

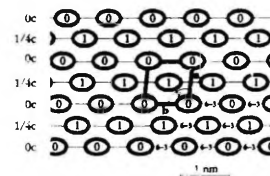


Fig. 8. Model P2. Two-chain unit cell. 01. Symbols: see Fig. 1 and 4.

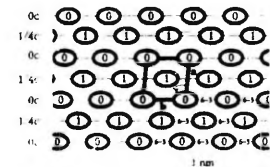


Fig. 9. Model P2. Two-chain unit cell. 01. Symbols: see Fig. 1 and 4.

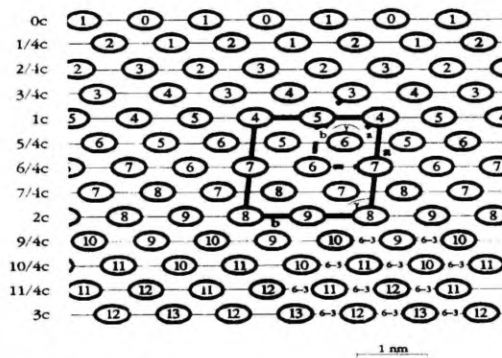


Fig. 10. Model A3a. Eight-chain unit cell: 01123243. Symbols: see Fig 1. and 4.

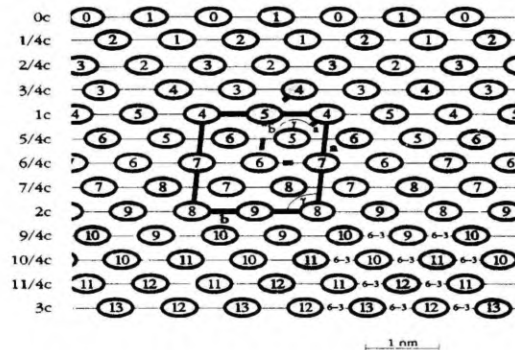


Fig. 11. Model A3b. Eight-chain unit cell: 01213234. Symbols: see Fig 1. and 4.

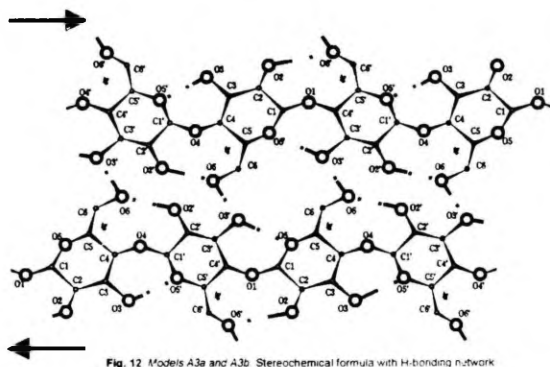


Fig. 12. Models A3a and A3b. Stereochemical formula with H-bonding network.

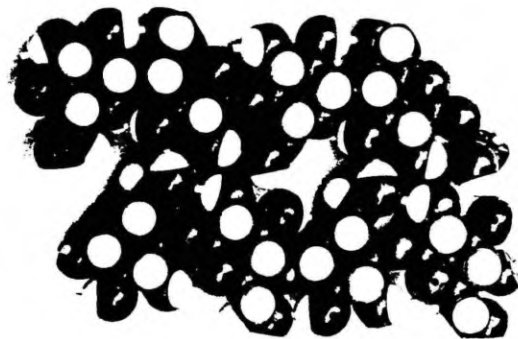


Fig. 13. Models A3a and A3b. Plastic space-filling molecular model. Cf Fig 12

Packing Energy Calculations on the Crystalline Structure of Cellulose I

Alvo Aabloo¹, Alfred D. French², Raik-Hiio Mikelsaar³ - ¹ Tartu University, Institute of Experimental Physics and Technology, Tähe 4 Street, EE2400 Tartu, Estonia; ² Southern Regional Research Center, USDA, 1100 Robert E. Lee Blvd, P.O. Box 19687, New Orleans, LA 70179, USA; ³ Tartu University, Institute of General and Molecular Pathology, Veski 34 Street, EE2400 Tartu, Estonia.

Abstract

The crystalline structure of native cellulose were investigated with packing energy calculations of crystals. Unit cell measures based on dimensions proposed by Sugiyama *et al* . The packing energy calculations of phases Ia and Ib of native cellulose (I) were evaluated using two different algorithms. As a first step we used rigid ring method. The best structures found by this method were further minimized with a full optimization molecular mechanics program MM3(90). Different minimization methods gave analogous structures for best models of both phases. The best model for phase Ia one-chain unit cell is in *tg* position and is packed "up". The chains of the best model for phase Ib two-chain unit cell are also in *tg* positions and are packed "up". The structure found for the phase Ib of native cellulose is completely similar with a structure proposed several years ago by Woodcock and Sarko for ramie cellulose.

Introduction

Several years ago, VanderHart and Atalla found that all crystalline native cellulose I is composed from two phases¹. Different samples consists these phases in different relation. Later Sugiyama *et al*², described those phases as Ia with a triclinic one-chain unit cell and as Ib with a monoclinic two chain unit cell. They investigated *Microdictyon tenuis* with electron diffraction techniques. The one-chain triclinic unit cell for the phase Ia has parameters of $a = 0.674$ nm, $b = 0.593$ nm, $c = 1.036$ nm, $\alpha = 117^\circ$, $\beta = 113^\circ$ and $\gamma = 81^\circ$. The two-chain unit monoclinic unit cell has parameters of $a = 0.801$ nm, $b = 0.817$ nm, $c = 1.036$ nm and $\gamma = 97.3^\circ$ as published by Sugiyama *et al*². Early these two phases together were indexed as eight-chain unit cell³.

Material and methods

We used two different strategies to find the most probable models using the crystal packing energy calculations. At the first step we used molecular modelling⁴ and rigid ring calculations^{5,6} to find the most probable models of both phases of native cellulose. The parallel packing of chains were presupposed. During these calculations we used $P1$ symmetry for the phase $I\alpha$ and $P2_1$ symmetry for the $I\beta$ phase. The rigid-ring method means that during the minimization process the glucose ring was kept fixed. The Arnott-Scott⁷ glucose ring was used. The torsion angles of the hydroxymethyl and hydroxyl groups and the torsion angles and the bond angles at the glycosidic linkages were allowed to vary (see Figure 1.). In the case of space group $P2_1$ the monomer residues were linked into the chain with virtual bond method⁸.

The best models of these calculations were further minimized with a full molecular mechanics program MM3(90)^{9,10}. During these minimizations the dielectric constant was chosen of 4. This value was also used for crystalline amides, polypeptides and proteins¹¹. In this method, a model of the crystal of cellulose was built of seven cellotetraose molecules placed to hexagonal close packing. All chains were arranged according to the unit cell measures.

The best model of both phases of cellulose I obtained by rigid-ring method preserved their leader position after MM3 minimization. While the conformation of these models have not changed significantly. The best model of both phases with needed results of calculations are presented in table 1. As we can see in the case of the phase $I\alpha$ the hydroxymethyl groups are in *tg* position. The orientation of chain is "up". In the case of the phase $I\beta$ the hydroxymethyl groups of both chains are also in *tg* position and they are also packed "up". The central chain is shifted down approximately $1/4$ c (see Figure 2.). Hydrogen bonds are completely same in the case of both phases. The conversion from $I\alpha$ phase to $I\beta$ phase is possible by simple vertical shifting of sheets of chains. The direct shifting of sheets needs to cross the barrier of 9 kcal/mol.

Results and discussion

The best model of the phase $I\alpha$ has an energy of -19.5 kcal/mol by rigid-ring method and 185 kcal/mol by MM3. The best model of the phase $I\beta$ has an energies of -19.9 kcal/mol and 182 kcal/mol, respectively. We can compare it with energies of -21.4 kcal/mol⁵ and 176 kcal/mol¹¹ calculated for cellulose II. Although the energies

Molecular Modelling of Parallel and Antiparallel Structure of Native Cellulose

Raili-Niilo Mikkeläseer and Alvo Aabloom^{*} - Tartu University, Institute of General and Molecular Pathology, 34 Väeski Street, EE2400 Tartu Estonia.
^{*}Tartu University, Institute of Experimental Physics and Technology, 4 Tähe Street, EE2400 Tartu, Estonia

Abstract

Tartu plastic space-filling atomic-molecular models were used to investigate the structure of cellulose crystalline (α and β) phases. It was elucidated that both parallel and antiparallel structures can be accommodated with unit cell geometries described by Sugiyama *et al.*¹. Packing energy calculations revealed that parallel models of α and β phases have similar energy. In contrast, in case of antiparallel structure models, β is energetically much more favourable in comparison with α phase. Our data indicate that antiparallel structure proposed for cellulose α phase is metastable and should be easily converted to β phase. So the antiparallel models are more suitable for explanations of cellulose $\alpha \rightarrow \beta \rightarrow \text{II}$ interconversion.

Introduction

Recently it was elucidated that native cellulose is a mixture of two crystalline phases (α and β) the proportion of which depends on the source of cellulose^{2,3}. Horii *et al.*⁴ showed that the α phase of cellulose is metastable and, through a hydrothermal annealing treatment, can be converted readily into the thermodynamically stable β phase. Sugiyama *et al.*¹ studied cellulose from a green alga *Microdictyon* by electron diffraction and found that the major α phase has a one-chain, triclinic structure and the minor β phase is characterized by two-chain unit cell and a monoclinic structure. This study disclosed the detail molecular structure of native cellulose crystalline phases. The aim of current paper was to use the molecular modelling method and packing energy calculations to elucidate the most favourable stereochemical structures corresponding to unit cell parameters found by Sugiyama *et al.*¹ in cellulose α and β crystalline phases.

Material and methods

Molecular modelling was carried out by Tartu plastic space-filling models having improved parameters and design⁵. Packing energy calculations were carried out by rigid-ring method⁶.

To present graphically the stereochemical models investigated, we drew unit cell projections on the plane crossing the c-axis at 90°. Cross-sections of up cellulose chains were given by white and down chains by grey ovals. The number of c/4 shifts of molecules in the direction of c-axis is noted by numbers inside of ovals⁷.

To make a short digital designation of the cellulose models: 1) symbols "P" and "A" were used to represent correspondingly "parallel" and "antiparallel", 2) numbers indicating c/4 shifts were applied for characterizing each chain in the unit cell, 3) underlining was used to differ up and down chains.

of both phases of native cellulose are very close, these energies are in a good agreement with experimental data about $\alpha \rightarrow \beta$ -II conversion. It is remarkable that the most probable model for $I\beta$ phase is completely analogous with a structure proposed by Woodcock and Sarko¹² for ramie cellulose. Also it is close to structure proposed by Pertsin *et al* for ramie cellulose using similar minimization technique¹³. It means that the diffraction patterns of ramie cellulose mostly belong to the phase $I\beta$. We got same result for phase $I\beta$ without any diffraction data.

Table 1. The best model for the phase $I\alpha$ and for the phase $I\beta$ of native cellulose.

| Model | Energy kcal/mol | | H-bonds in chain | H-bonds in sheet | -CH ₂ OH position |
|-----------|-----------------|-----|------------------|------------------|------------------------------|
| | Rigid-ring | MM3 | | | |
| $I\alpha$ | -19.5 | 185 | O5..HO3 O6..HO2 | O3..HO6 | <i>tg</i> |
| $I\beta$ | -19.9 | 182 | O5..HO3 O6..HO2 | O3..HO6 | <i>tg</i> |

References

1. VanderHart D. L.; Atalla R. H., *Macromolecules* 1984, v. 17, p. 1465.
2. Sugiyama, J.; Vuong, R; Chanzy, H. *Macromolecules* 1991, v. 24, p. 4168.
3. Honjo G.; Watanabe M. *Nature* 1958, v. 181, p. 326.
4. Mikelsaar, R. *Trends Biotechnol.* 1986, v. 6, p. 162.
5. Pertsin, A. J.; Nugmanov, O. K.; Marchenko, G. N.; Kitaigorodsky, A. I. *Polymer* 1984, v. 25, p. 107.
6. Mikelsaar, R.-H.; Aabloo, A. In: *Cellulosics: Chemical, Biochemical and Material Aspects*. Eds. J. F. Kennedy, G. O. Phillips, P. A. Williams. Ellis Horwood, Chichester, 1993, p. 61.
7. Arnott S.; Scott W. E. *J. Chem. Soc. Perkin II* 1972, p. 324.
8. Zugenmayer P.; Sarko A. *ACS series* 1980, v. 141, p. 226.
9. Allinger N. L.; Yuh Y. H.; Lii J.-H. *J. Amer. chem. Soc.* 1989, v. 111, p. 8551.
10. Allinger N. L.; Rahman, M.; Lii J.-H. *J. Amer. chem. Soc.* 1990, v. 112, p. 8293.
11. Stipanovic A. J.; Sarko A. *Macromolecules* 1976, v. 9, p. 851.
12. Woodcock C.; Sarko A. *Macromolecules* 1980, v. 13, p. 1183.
13. Pertsin A. J.; Nugmanov O. K.; Marchenko G. N. *Polymer* 1986, v. 27, p. 597.

Results

The characteristics of the cellulose structure models investigated is presented in Table 1.

1. Molecular models fitting with cellulose Ia unit cell parameters

1.1. Models P1

The investigation of green alga *Microdictyon* by Sugiyama *et al.*¹ is the first work where unit cell parameters are given for pure cellulose Ia phase microcrystals: $a = 0.674$ nm, $b = 0.593$ nm, $c = 1.036$ nm, $\alpha = 117^\circ$, $\beta = 113^\circ$ and $\gamma = 81^\circ$. This one-chain unit cell requires parallel packing of molecular chains. Molecular modelling by Tartu devices showed that if there are the usual intramolecular bonds O3-H...O5 and O2-H...O6 the hydroxymethyl groups should be in *tg* conformation and all parallel chains are bound by intrasheet bonds O6-H...O3. Packing energy calculations of the above-described cellulose structure imitation revealed that model P1 has potential energy -19.5 kcal/mol.

1.2. Models A1a and A1b

Two antiparallel models A1a and A1b characterized by eight-chain unit cell maybe geometrically derived from the one-chain unit cell described by Sugiyama *et al.*¹. The geometry of these two models is very similar. The only difference is that in A1a parallel chains follow each other from down right to up left dividing major angles of unit cell and in A1b from down left to up right dividing minor angles of unit cell. So A1a may be called "left" and A1b "right" type of structures. According to molecular modelling data chains having usual intramolecular H-bonds and *tg* conformation of hydroxymethyl groups should be in models A1a and A1b bound by intrasheet bonds O6-H...O2. The potential energy values of above-described structures are rather similar: model A1a has -16.4 kcal/mol and A1b -15.5 kcal/mol.

2. Molecular models fitting with cellulose Ib unit cell parameters

2.1. Model P2

Sugiyama *et al.*¹ have established that Ib phase of *Microdictyon* cellulose has two-chain monoclinic structure with unit cell parameters of $a = 0.801$ nm, $b = 0.817$ nm, $c = 1.036$ nm and $\gamma = 97.3^\circ$. The authors suppose that this structure consists of parallel molecules and the center and corner chain are staggered by $c/4$. Molecular modelling revealed that intra- and intermolecular H-bonds in model P2 are identical to those in model P1. The energy value of model P2 is close to model P1: -19.9 kcal/mol.

2.2. Model A2

Molecular modelling showed that a chain conformation and an intrasheet H-bonding network in model A2 are identical to model P2. The only difference is that sheets are parallel in model P2 and antiparallel in model A2. The potential energy of model A2 is relatively high: -15.8 kcal/mol.

2.3. Models A3a and A3b

Two antiparallel models A3a and A3b characterized by eight-chain unit cell may be geometrically derived from the two-chain unit cell described by Sugiyama *et al.*¹. These structures are similar to the

"left" and "right" models A1a and A1b differing mainly by presence of intrasheet stagger of cellulose chains. The investigation by plastic models indicated that in the case where standard intramolecular H-bonds and tg conformation of hydroxymethyl groups are preserved the antiparallel molecular chains are bound by bonds O6-H...O3. There is considerable difference between energies of above-described two structures: the "left" type model A3a has -21.0 kcal/mol and "right" model A3b -15.1 kcal/mol.

Discussion

Our investigation revealed that both parallel and antiparallel structure models can be proposed based on cellulose I α and I β crystalline phases unit cell parameters published by Sugiyama *et al.*¹. Although they all have usual chain conformation and intramolecular H-bonding network, the potential energy values between seemingly close structures in certain cases are very different.

Sugiyama *et al.*¹ interpret their experimental data in the light of all-parallel structure view. Following this view, the I α phase structure should correspond to our model P1 and I β phase to model P2.

Comparing cellulose parallel model P1 with antiparallel models A1a and A1b one can see that these models are sterically very similar. If the H-bonding network and chain polarity are not clearly established the structure characterized by eight-chain unit cell can be considered as the one-chain unit cell. We think that antiparallel model A1a having lower energy (-16.4 kcal/mol) in comparison with the model A1b (-15.5 kcal/mol) is a possible pretender for a "masked" eight-chain unit cell in the cellulose I α crystal structure described by Sugiyama *et al.*¹. The antiparallel structure is not excluded also for cellulose I β crystalline phase because these authors marked that the exact position of the cellulose chains in the monoclinic unit cell was even more difficult to ascertain as long as the structure refinement was not achieved. No doubt, A3a is the best antiparallel model for fitting to the unit cell structure of I β crystalline phase because it has an excellent packing energy: -21.0 kcal/mol.

Table 1. Parallel and antiparallel stereochemical models for cellulose I α and I β crystalline phases.

| Phase | Model | Unit cell designation | Intramolecular H-bonds | Intrasheet H-bonds | -CH ₂ OH group conformation | Packing energy kcal/mol |
|------------|-------|-----------------------|------------------------|--------------------|--|-------------------------|
| I α | P1 | 1 | O3-H...O5 O2-H...O6 | O6-H...O3 | tg | -19.5 |
| I β | P2 | 01 | O3-H...O5 O2-H...O6 | O6-H...O3 | tg | -19.9 |
| I α | A1a | 00112233 | O3-H...O5 O2-H...O6 | O6-H...O2 | tg | -16.4 |
| I α | A1b | 00112233 | O3-H...O5 O2-H...O6 | O6-H...O2 | tg | -15.5 |
| I β | A2 | 01 | O3-H...O5 O2-H...O6 | O6-H...O3 | tg | -15.8 |
| I β | A3a | 01123243 | O3-H...O5 O2-H...O6 | O6-H...O3 | tg | -21.0 |
| I β | A3b | 01213234 | O3-H...O5 O2-H...O6 | O6-H...O3 | tg | -15.1 |

If the "left" type antiparallel models A1a and A3a are thought to be the candidates for structures of cellulose I α and I β crystalline phases, correspondingly, it would be easy to explain the experimental data on the interconversion of these phases. Model A1a, having relatively high energy -16.4 kcal/mol, is metastable and can during annealing process easily be converted irreversibly into the model A3a, having very low energy -21.0 kcal/mol. Because the models A1a and A3a have different H-bonding networks, our conception on antiparallel cellulose structure is in accordance also with Raman-

spectroscopic data⁸. The high energy barrier between I α and I β crystalline phases, bound with rearrangement of hydrogen bonds, may be the reason why the proportion of these phases in different samples of native cellulose is relatively constant. The notable structural differences between models A1a and A3a allow one also to understand the differences between corresponding unit cell parameters of I α and I β cellulose crystalline phases.

The process of cellulose I \rightarrow II conversion can be also easily interpreted by all-antiparallel conception, because already Na-cellulose I, formed in the initial step of mercerization, has antiparallel structure^{9,10}.

References

1. Sugiyama, J.; Vuong, R.; Chanzy, H. *Macromolecules* 1991, v. 24, p. 4168.
2. Atalla, R. H.; VanderHart, D. L. *Science* 1984, v. 223, p. 283.
3. VanderHart, D. L.; Atalla, R. H. *Macromolecules* 1984, v. 17, p. 1465.
4. Horii, F.; Yamamoto, H.; Kitamaru, R.; Takahashi, M.; Higuchi, T. *Macromolecules* 1987, v. 20, p. 2946.
5. Mikelsaar, R. *Trends Biotechnol.* 1986, v. 6, p. 162.
6. Pertsin, A. J.; Nugmanov, O. K.; Marchenko, G. N.; Kitaigorodsky, A. I. *Polymer* 1984, v. 25, p. 107.
7. Mikelsaar, R.-H.; Aabloo, A. In: *Cellulosics: Chemical, Biochemical and Material Aspects*. Eds. J. F. Kennedy, G. O. Phillips, P. A. Williams. Ellis Horwood, Chichester, 1993, p. 57.
8. Atalla, R. H. In: *Cellulose Structural and Functional Aspects*. Eds. J. F. Kennedy, G. O. Phillips, P. A. Williams. Ellis Horwood, Chichester, 1989, p.61.
9. Sarko, A.; Nishimura, H.; Okano, T. In: *Structure of Cellulose*; ACS Symposium Series 340; American Chemical Society: Washington, DC, 1987 p. 169.
10. Nishimura, H.; Okano, T.; Sarko, A. *Macromolecules* 1991, v. 24, p. 759.

